Interference Management in a Class of Multi User Networks

by

Seyyed Hassan Mahboubi

A thesis

presented to the University of Waterloo

In fulfillment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2014

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Seyyed Hassan Mahboubi

Abstract

Spectrum sharing is known as a key solution to accommodate the increasing number of users and the growing demand for throughput in wireless networks. Interference is the primary barrier to enhancing the overall throughput of the network, especially in the medium and high signal to noise ratios (SNRs). Managing interference to overcome this barrier has emerged as a crucial step in developing efficient wireless networks.

An interference management strategy, named interference Alignment, is investigated. It is observed that a single strategy is not able to achieve the maximum throughput in all possible scenarios, and in fact, a careful design is required to fully exploit all available resources in each realization of the system.

In this dissertation, the impact of interference on the capacity of X networks with multiple antennas is investigated. Degrees of freedom (DoF) are used as a figure of merit to evaluate the performance improvement due to the interference management schemes. A new interference alignment technique called layered interference alignment, which enjoys the combined benefits of both vector and real alignment is introduced in this thesis. This technique, which uses a type of Diophantine approximation theorems first introduced by the author, is deployed and was proved to enable the possibility of joint decoding among the antennas of a receiver. With a careful transmitter signal design, this method characterizes the total DoF of multiple-input multiple-output (MIMO) X channels. Then, this result is used to determine the total DoF of two families of MIMO X channels. The Diophantine approximation theorem is also extended to the field of complex numbers to accommodate the complex channel realizations as well.

Acknowledgements

I include amongst my blessings the presence of many individuals whose help and support was priceless in my academic achievements and career.

First and foremost, I would like to express my sincere gratitude to my supervisor, Dr. Amir K. Khandani, for his inspiration and consistent guidance throughout the course of my studies at the University of Waterloo. I also wish to thank the members of my dissertation committee, Professors Abbas Yongacoglu, Liang-Liang Xie, Oleg Michailovich, and Shoja'eddin Chenouri for having agreed to take the time out of their busy schedules to read my thesis and to provide me with their comments and suggestions.

I would like to specially acknowledge Dr. Mumtaz Hussein for his help and contributions in the proof of the number theoretic part of my results; without his help this research might have taken much longer. I would like to thank all Coding and Signal Transmission lab members, as they created a very pleasant, intellectually stimulating, and friendly environment. In particular, I would like to thank Dr. Abolfazl Motahari, Dr. Akbar Ghasemi, and Ehsan Ebrahimzad with whom I have enjoyed sharing ideas and discussions during the course of my Ph.D. studies. My warm wishes and thanks go as well to my dear friends Dr. Seyed Ali Ahmadzadeh, Dr. Moein Mehrtash, Dr. Alireza Bayesteh, Javad, Saman, Mohsen, Hojat, Babak, Ershad, and many others who have enriched my life beyond my studies.

Last but not the least, I would like to express my warmest gratitude to my wife, Dr. Golmehr Sistani, for her endless support, patience, and encouragement, without which I would not have finished this long journey of Ph.D. studies. In addition, I would like to thank my parents, my grandparents, and my sister for their continuous support and encouragement throughout my life.

To my beloved wife, Dr. Golmehr Sistani,

and

My parents, Maryam & Naser

Contents

List of Figures

List of Abbreviations

AWGN	Additive White Gaussian Noise
BC	Broadcast Channel
DoF	Degrees Of Freedom
GIC	Gaussian Interference Channel
IC	Interference Channel
i.i.d.	independent and identically distributed
MAC	Multiple Access Channel
MIMO	Multiple-Input Multiple-Output
SIMO	Single-Input Multiple-Output
SISO	Single-Input Single-Output
SNR	Signal to Noise Ratio

Notation and List of Symbols

Throughout this thesis, boldface upper-case letters, e.g., $\mathbf{H}$, are used to represent matrices. Matrix elements will be shown in brackets throughout this article, e.g., $\mathbf{H} = [h_{i,j}]$ for a set of values i, j. Vectors are shown using boldface italic lower-case letters, e.g., $\boldsymbol{v}$. Vector elements are shown inside parenthesis, e.g., $\boldsymbol{v} = (v_1, v_2, ..., v_i)$ for a set of values i. The transpose and conjugate transpose of a matrix $\mathbf{A}$ will be represented as $\mathbf{A}^t$ and $\mathbf{A}^\dagger$. The general transmitted signal from the kth antenna of transmitter i desired to be decoded at receiver j is represented by $x_k^{i,j}$. At each antenna of transmitter in the X channel, a linear combination of all desired messages for different receivers will be transmitted. To simplify notation it is assumed that $x_k^i = \sum_j \beta_j x_k^{i,j}$, where β_j is the weight of message $x_k^{i,j}$ in the linear combination. The transmitted vector signal at transmitter i will be represented as $\boldsymbol{x}^i = (x_1^i, x_2^i, ..., x_k^i)^t$ for a set of values k. We use single superscript labelling for the indexes of transmitters and receivers, for example, $\boldsymbol{z}^i$ represents the noise vector at the receiver i. Single subscripts are used for the antenna labelling, unless otherwise stated; for example, y_j^i represents the received signal at the jth antenna of receiver i. The superscript pair i, j assigns the variable from transmitter i to receiver j, and similarly the subscript pair l, n represents the variable from antenna l to antenna n. For example, $h_{l,n}^{i,j}$ represents the channel between the lth antenna of transmitter i and the nth antenna of the receiver j. Other notations are listed below:

$\|\mathbf{A}\|$	Determinant of matrix $\mathbf{A}$
$\mathbf{A} \geq 0$	Matrix A is positive semi-definite
K	Number of users or receivers
M	Number of antennas at users or receivers
$\mathbb{R}$	The set of real numbers
$\mathbb{R}^n$	The n-dimensional Euclidean space
$\mathbf{Q}$	The set of rational numbers
$\mathbb{N}$	The set of natural numbers
$\mathbb{Z}$	The set of all integers
$E[X]$	The expectation of the random variable X
$\mathbb{C}$	The set of complex numbers
$\|U\|$	The cardinality of a set U
$(a,b)_{\mathbb{Z}}$	The set of integers between a and b.

Chapter 1

Introduction

The study of interaction between non-cooperative users sharing the same channel goes back to Shannon's work on the two-way channel in [1]. Several researchers followed his work, and the two-user interference channel (IC) emerged as a fundamental problem regarding interaction between users causing interference in the networks. In this channel, two senders transmit independent messages to their corresponding receivers via a common channel. The characterization of the channel's capacity region that reveals the acceptable rates in the system has been an open problem for more than 40 years.

There are some special cases where the exact capacity region has been characterized. A limiting expression for the capacity region was obtained in [2] (see also [3]). Owing to excessive computational complexity, this expression cannot be used directly to fully characterize the capacity region. To show this, Cheng and Verdú [4] proved that for the Gaussian multiple access channel (MAC), which can be considered as a special case of the Gaussian interference channel (GIC),[1] the limiting expression fails to fully characterize the capacity region by relying only on the Gaussian distributions . There are, however, some special cases where the limiting expressions can be optimized. For example, the sum capacity of the Gaussian MAC can be achieved by relying on the simple scheme of frequency/time division multiple access [5].

During the past three decades, information theorists have made extensive efforts to charac-

[1]In a special case of the GIC, the received signals at both receivers are statistically equivalent. The capacity region of this channel is equivalent to that of the Gaussian MAC observed from one of the receivers.

terize the impact of the interference on the capacity of wireless networks. The GIC is a simple model that captures most of the characteristics of wireless networks. This model describes a communication system with multiple transmitter/receiver pairs in which each transmitter wishes to communicate with its corresponding receiver while generating interference to all other receivers. Although there is an extensive body of research on the capacity characterization of the GIC, the capacity region of this channel is still generally unknown even for the simplest case of a two-user GIC. In fact, for the two-user GIC, the capacity region has been completely characterized only for some ranges of channel coefficients [6, 7, 8, 9, 10, 11, 12]. For the general two-user case, a characterization of the capacity region within one bit has been presented in [13].

Moving from the two-user case to more than two users makes the capacity characterization more challenging. To reduce the severe effect of the interference for the case of $K > 2$ users, the use of a new technique known as *Interference Alignment* is essential. A version of interference alignment, which was first introduced by Maddah-Ali et al. [14, 15], is an elegant technique that reduces the effect of the aggregated interference from several users to the effect of the interference from only one user. In this technique a subspace is dedicated to interference, and the signaling is designed such that all the interfering signals are squeezed into the interference subspace [15, 16]. There are two versions of interference alignment in the literature: signal space alignment and signal scale alignment. In the signal space alignment method, careful choosing of transmission directions converts the interference channel into multiple non-interfering channels. Signal space alignment approaches are applicable to interference channels with time varying/frequency selective channel coefficients. Signal scale alignment schemes, on the other hand, use structured coding, e.g., lattice codes, to align interference at the signal level and are especially useful in the case of static channels. In [17], for the special cases of many-to-one and one-to-many GICs, the authors have computed the capacity region within constant bits using the signal scale alignment technique. For the general case of $K > 2$ users most of the effort has focused on the characterization of the total degrees of freedom. The total degrees of freedom (DoF) for a GIC shows the growth of the maximum achievable sum rate in the limit of increasing Signal to Noise Ratio (SNR). Using the idea of signal space interference alignment, Cadambe and Jafar [18] showed that for a fully connected K-user GIC with time varying/frequency selective channel coefficients, the total degrees of freedom is equal to $\frac{K}{2}$. Since interference alignment needs full channel state information at all transmitters, the assumption of availability of channel gains for time-varying channels is far from being practical, and hence the results of Cadambe and Jafar cannot be applied to the more realistic case of static channels.

Etkin and Ordentlich [19] deployed some results of additive combinatorics to show that for a constant fully connected real GIC, the total degrees of freedom is very sensitive to the rationality/irrationality of channel coefficients. They showed that for a K-user fully connected constant real GIC with rational channel coefficients, the total degrees of freedom is strictly less than $\frac{K}{2}$. Moreover, they showed that for a class of measure zero of channel coefficients, the total degrees of freedom of this channel is equal to $\frac{K}{2}$. Motahari et al. [16] showed that by deploying a new type of interference alignment, which the authors have called real interference alignment, it is possible to achieve $\frac{K}{2}$ degrees of freedom for almost all K-user real GICs with constant channel coefficients. The essence of this method is to align discrete points along a real axis based on the number theoretic properties of rational and irrational numbers [16].

The basic idea of *Vector Interference Alignment* initiated from the 2×2 X channel with three antennas at each receiver. Later on, for the setup of 2×2 multiple-input multiple-output (MIMO) X channels, a technique called symbol extension [20] is introduced in the context of time varying MIMO X channels. Motahari et al. [21] used the real interference alignment over the single-input single-output (SISO) X channels with constant channel realizations. It is noteworthy that although initially the implementation of interference alignment looks simpler over X channel than GIC, when the number of transmitters/receivers increases, the problem will be much more sophisticated. There are efforts to characterize the total DoF of the time varying/frequency selective X channels in the literature; most of these rely on symbol extension and MIMO techniques like *zero-forcing* and *beam-forming*.

In this thesis, for the first time an attempt to characterize the total DoF of MIMO X channels with constant channel realization is discussed. A new mathematical tool is used to find a new number theoretic theorem, which can provide required tools for characterization of the total

DoF of a class of MIMO X channels. This new Diophantine approximation theorem enables the power of multiple antenna diversity along with the possibility of modulating the transmit signal using the rational dimensions. At the receivers, by using this new theorem, one can exploit the simultaneous/joint decoding at all antennas. This technique can maximize the total fractional achievable DoF. For the first time, in this dissertation the complex channel realization is also considered. It is assumed that each channel gain can be modeled with a complex number consisting of an amplitude and a phase. The achievability scheme for the complex channel gains is backed with a new complex number Khintchin–Groshev type of theorem that is introduced and proved in the last chapter of this thesis.

The rest of this thesis is organized as follows. In Chapter 2, some background and required preliminaries on the characterization of DoF for X channels are reviewed. I will focus on the concepts of the real interference alignment technique in achieving the DOF of SISO X channels.

In Chapter 3, the required mathematical tools and techniques for introducing and proving the joint decoding theorem are discussed in detail. After the provision of the required analytical tools, the insight to a new number theoretic theorem, which will be introduced and proved in the next chapter, is discussed.

In Chapter 4, results of Chapter 3 are used to characterize the total DoF of both $K\times 2$ and $2\times K$ MIMO X channels with constant real channel realizations. The required encoder and decoder design for each case is provided, and it is observed that the total DoF is achieving the outerbound and is the same as the time varying/frequency selective channel with a much more complicated signaling. This combination of signaling and joint decoding is called *Layered Interference Alignment*. The detailed mathematical proof for the new Diophantine approximation theory of the layered interference alignment is provided. To extend these results to the complex channel realizations a new set of tools is required. The Diophantine approximation bound provided for the complex numbers is totally different from the real number case, but analysis for the total DoF per transmit dimension is retained in comparison with the real channel realizations.
Finally, the conclusion and potential research pathway are presented in Chapter 5.

Chapter 2

Background and Preliminaries

Sharing the available wireless medium for higher data transmission has made interference management one of the biggest challenges in wireless networks. Spectrum sharing is known as a key solution to time/frequency allocation among users. However, in dense networks achieving the optimum throughput of the system is not obtained only by orthogonal schemes, making interference management inevitable. Extensive efforts have been made by information theorists to characterize the ultimate obstruction that interference imposes on the capacity of wireless networks. In order to reduce the severe effect of interference for the $K > 2$ users interference channel, the use of a new technique known as interference alignment is crucial.

Interference alignment was first introduced by Maddah-Ali et al. [15] in the context of multiple-input multiple-output (MIMO) X channels as a breakthrough technique that makes the interference less damaging by merging the communication dimensions occupied by interfering signals.

Interference alignment in n-dimensional Euclidean spaces for $n \geq 2$, known as vector interference alignment has been studied by several researchers, e.g., [15, 20, 22, 23]. In this method, at each receiver a subspace is dedicated to interference; then the signaling is designed such that all the interfering signals are squeezed into the interference subspace. Using this method, Cadambe and Jafar [22] showed that, contrary to popular belief, a K-user Gaussian interference channel (GIC) with varying channel gains could achieve its total DoF, which is $\frac{K}{2}$. Since the assumption of varying channel gains is unrealistic, particularly that all the gains should be known at the transmitters, the application of these important theoretical results is limited in practice.

Motahari et al. [24] settled the problem for the general scenario by proposing a new type of interference alignment that can achieve $\frac{K}{2}$ DoF for almost all K-user real GIC with constant coefficients. This result was obtained by introducing a new type of interference alignment known as real interference alignment. In this technique, tools from the field of Diophantine approximation in number theory play a crucial role [25].

Extending the aforementioned results to the K-user MIMO interference channel is straightforward when the number of transmitter antennas is equal to the number of receiver antennas. Also, studies like [22, 24] showed that for a K-user M-antenna MIMO interference channel the total number of DoF is equal to $\frac{KM}{2}$, whether the channel is constant or time varying/frequency selective. However, extension to the general K-user, with M antennas at transmitter and N antennas at receiver, interference channel is not straightforward. Ghasemi et al. [26] partially settled the problem by proposing tighter upper and lower bounds for the MIMO constant channels.

The MIMO X channel behaves differently compared with the MIMO K-user GIC. Although in the latter the total DoF is fully characterized for an equal number of antennas at all nodes, the corresponding problem in the former setup is still open. This is because vector or real interference alignment techniques cannot provide the necessary means to settle the problem individually. Mahboubi et al. [27] were the first to introduce a new type of alignment technique, called layered interference alignment, in which real interference alignment is used in conjunction with vector alignment to obtain optimal signaling for the MIMO $K \times 2$, X channel. The layered interference alignment not only enjoys the benefits of both its ancestor techniques but also relies on a new Khintchine–Groshev type inequality that is introduced for the first time in this dissertation. This theorem enables the receiver to exploit the availability of multiple antennas at its side. In this chapter the details of the real interference alignment technique and how it is used to characterize the total DoF of single antenna 2×2 and $K\times M$ X channels are explained.

2.1 Degrees of Freedom

Besides the capacity results, there are a number of works on characterizing the degrees of freedom of the vector GIC. The degrees of freedom of a point-to-point communication system is the capacity pre-log factor for large values of signal to noise ratio (SNR). More precisely, if the capacity of the system can be formulated as $C(\text{SNR}) = d\log_2(\text{SNR}) + o(\log_2(\text{SNR}))$ [1] then the system has d degrees of freedom. This definition is for complex settings (all signals, noise, and channel coefficients are complex variables). In the real setting, it is said that a point-to-point system has d degrees of freedom if the capacity of the system can be expressed as $C(\text{SNR}) = \frac{d}{2}\log_2(\text{SNR}) + o(\log_2(\text{SNR}))$. Note that with this definition the degrees of freedom of the system would be independent of the underlying real or complex setting.

According to this definition, a point-to-point single-input single-output (SISO) Gaussian channel has one degree of freedom. The point-to-point MIMO channel with M transmit antennas and N receive antennas is known to have $\min(M, N)$ degrees of freedom [28]. Although this result can be obtained from the capacity results of this channel, it can also be obtained by using simpler methods like zero-forcing.

For multi-user channels a degrees of freedom region (similar to capacity region) can be defined. The degrees of freedom region of a channel is in fact the shape of the capacity region of that channel in the high SNR regime scaled by $\log_2(\text{SNR})$. Let $\mathscr{C}$ and $\mathscr{D}$ denote the capacity region and degrees of freedom region of a multi-user channel, respectively. All extreme points of $\mathscr{D}$ can be identified by solving the following optimization problem

$$D = \begin{cases} \lim_{\text{SNR}\to\infty} \frac{\max_{\mathbf{R}\in\mathscr{C}} \boldsymbol{\omega}^t\mathbf{R}}{\log_2(\text{SNR})} & \text{for complex setting} \\ \\ \lim_{\text{SNR}\to\infty} \frac{\max_{\mathbf{R}\in\mathscr{C}} \boldsymbol{\omega}^t\mathbf{R}}{0.5\log_2(\text{SNR})} & \text{for real setting} \end{cases} \tag{2.1}$$

here ω is the connectivity vector. Then the individual degrees of freedom of user i are defined

[1] $f(n) = o(g(n))$ means $f(n)$ becomes insignificant relative to $g(n)$ as n approaches infinity. More precisely $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$

as

$$d^i = \begin{cases} \lim_{\text{SNR}\to\infty} \frac{R^i}{\log_2(\text{SNR})} & \text{for complex setting} \\ \\ \lim_{\text{SNR}\to\infty} \frac{R^i}{0.5\log_2(\text{SNR})} & \text{for real setting} \end{cases} \tag{2.2}$$

where R^i is any allowable rate for user i within the capacity region.

Similar to the sum-rate capacity, the total degrees of freedom of a multi-user system, D, with K users refers to the case of $\omega = (1, 1, \cdots, 1)^t$ in (2.1)

$$D = \max_{(d^1,\cdots,d^k)\in\mathcal{D}} (d_1 + d_2 + \cdots + d_K)$$

In fact, the total degrees of freedom is the sum-rate capacity in a high SNR regime scaled by $\log_2(\text{SNR})$. For a two-user MIMO Gaussian MAC with M_1 and M_2 antennas at the transmitters and N antennas at the receiver, Jafar and Fakhereddin [29] proved that the total degrees of freedom is given by

$$D = \min(M_1 + M_2, N)$$

This means that the total degrees of freedom of the two-user MIMO MAC are the same as when there is perfect cooperation between the users. Similarly, for a MIMO Gaussian broadcast channel (BC) with M antennas at the transmitter and N_1 and N_2 antennas at the receivers, the total degrees of freedom can be expressed as follows [29]:

$$D = \min(M, N_1 + N_2)$$

Finally, for a MIMO GIC with M_1 and M_2 antennas at the transmitters and N_1 and N_2 antennas at their corresponding receivers, the total degrees of freedom is given by [29]

$$D = \min(M_1 + M_2, N_1 + N_2, \max(M_1, N_2), \max(M_2, N_1))$$

This result is obtained by a careful selection of transmit directions and by performing zero-forcing at the receivers. As special cases, the total degrees of freedom of a two-user scalar GIC is equal to 1 and the total degrees of freedom of a two-user vector GIC with M antennas at each terminal is equal to M.

2.1.1 Interference Alignment

In contrast to the major research activities on the interference channels, e.g., [8, 9, 13, 30], the problem of characterizing the capacity region of GIC is still open. It is shown by Motahari et al. [8] that in the two-user GIC, the Han–Kobayashi scheme [31] achieves within one bit of the capacity region, as long as the interference from the private message in the Han–Kobayashi scheme is designed to be below the noise level.

The result of [13] has provided a clear understanding about the behaviour of the two-user GIC. However, it turns out that moving from the two-user scenario to a larger number of users is a challenging task. Indeed, for a K-user GIC ($K > 2$), the Han–Kobayashi approach of managing the interference is not enough, and a new approach of interference management known as interference alignment must be incorporated.

Interference alignment is a solution for making the interference less severe at receivers by merging the communication dimensions occupied by the interfering signal. Maddah-Ali et al. [15] introduced the concept of interference alignment and showed its capability in achieving the full degrees of freedom for certain classes of two-user X channels. Because it is simple and at the same time powerful, interference alignment provided the spur for further research. It is not only usable for lowering the harmful effect of the interference but also can be applied to provide security in networks as proposed in [32].

Several researchers [15, 18, 20, 23] studied the interference alignment in n-dimensional Euclidean spaces for $n \geq 2$. In this method, at each receiver a subspace is dedicated to interference; then the signaling is designed such that all the interfering signals are squeezed into the interference sub-space. Such an approach saves some dimensions for communicating desired signals, rather than wasting them because of the interference. Using this method, Cadambe and Jafar [18] showed that, contrary to popular belief, a K-user Gaussian interference channel with varying channel gains could achieve its total DoF, which is $\frac{K}{2}$. Later, in Nazar et al. [33], it is shown that the same result can be achieved using a simple approach based on a particular pairing of the channel matrices. The assumption of varying channel gains, particularly noting that all the gains should be known at the transmitters' sides, is unrealistic, which limits the application of these important theoretical results in practice.

In [17], followed by [34, 35], the application of interference alignment is extended from two or more spatial/temporal/frequency dimensions to one dimension, but at the signal level. In [17] it is shown that lattice codes, rather than random Gaussian codes, are essential parts of signaling for three-user time-invariant GICs. In [34], after interference was aligned using lattice codes, the aggregated signal is decoded and its effect is subtracted from the received signal. In fact, Sridharan et al. [34] show that the very strong interference region of the K-user GIC is strictly larger than the corresponding region when alignment is not applied. In their scheme, to make the interference less severe, transmitters use lattice codes to reduce the code-rate of the interference, which guarantees decodability of the interference at the receiver. Sridharan et al. [35] showed that the DoF of a class of three-user GICs with fixed channel gains could be greater than one. This result was obtained using layered lattice codes along with successive decoding at the receiver.

In Etkin and Ordentlich [36] and Motahari et al. [37], the results from the field of Diophantine approximation in number theory are used to show that interference can be aligned using properties of rational and irrational numbers and their relations. These authors showed that the total DoF of some classes of time-invariant single antenna interference channels could be achieved. In particular, Etkin and Ordentlich [36] proposed an outerbound on the total DoF that maintains the properties of channel gains with respect to being rational or irrational. Using this outerbound, surprisingly they proved that the DoF is everywhere discontinuous for the class of channels under investigation.

2.2 MIMO X Channel

A 2×2 MIMO X channel is a system with two transmitters and two receivers, each equipped with multiple antennas, where independent messages need to be conveyed from each transmitter to each receiver. In this thesis, the benefits of transmitter side cooperation in the form of shared messages that are available to both transmitters is considered.

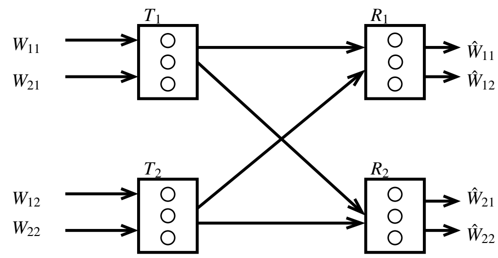

Figure 2.1: MIMO X Channel

The MIMO X channel is shown in Figure 2.1 and is described by the input output equations:

$$\begin{aligned} \boldsymbol{y}^1 &= \mathbf{H}^{1,1}\boldsymbol{x}^1 + \mathbf{H}^{2,1}\boldsymbol{x}^2 + \boldsymbol{z}^1 \\ \boldsymbol{y}^2 &= \mathbf{H}^{1,2}\boldsymbol{x}^1 + \mathbf{H}^{2,2}\boldsymbol{x}^2 + \boldsymbol{z}^2 \end{aligned}$$

where $\boldsymbol{y}^1$ is the $N_1 \times 1$ output vector at receiver 1, $\boldsymbol{y}^2$ is the $N_2 \times 1$ output vector at receiver 2, $\boldsymbol{z}^1$ is the $N_1 \times 1$ additive white Gaussian noise (AWGN) vector at receiver 1, $\boldsymbol{z}^2$ is the $N_2 \times 1$ AWGN vector at receiver 2, $\boldsymbol{x}^1$ is the $M_1 \times 1$ input vector at transmitter 1, $\boldsymbol{x}^2$ is the $M_2 \times 1$ input vector at transmitter 2, $\mathbf{H}^{1,1}$ is the $N_1 \times M_1$ channel matrix between transmitter 1 and receiver 1, $\mathbf{H}^{2,2}$ is the $N_2 \times M_2$ channel matrix between transmitter 2 and receiver 2, $\mathbf{H}^{2,1}$ is the $N_2 \times M_1$ channel matrix between transmitter 2 and receiver 1, and $\mathbf{H}^{1,2}$ is the $N_1 \times M_2$channel matrix between transmitter 1 and receiver 2. As shown in Figure 2.1, there are four independent messages in the MIMO X channel, $W_{11}, W_{12}, W_{21}, W_{22}$, where W_{ij} represents a message from transmitter j to receiver i. The power at each transmitter is assumed to be equal to ρ. The size of the message set is indicated by $|W_{ij}(\rho)|$. For codewords spanning n channel uses, the rates $R^{i,j}(\rho) = \frac{\log |W_{ij}(\rho)|}{n}$ are achievable if the probability of error for *all* messages can be simultaneously made arbitrarily small by choosing an appropriately large n. For rate functions $R^{ij}(\rho)$ the degrees of freedom are defined as

$$d^{i,j} = \lim_{\rho\to\infty} \frac{R^{i,j}(\rho)}{\log(\rho)}. \tag{2.3}$$

The degrees of freedom region for the MIMO X channel is defined as

$$\mathscr{D} \triangleq \left\{(d^{1,1}, d^{1,2}, d^{2,1}, d^{2,2}) : d^{i,j} = \lim_{\rho\to\infty} \frac{R^{i,j}(\rho)}{\log(\rho)},\ \ \Pr(\hat{W}_{ij} \neq W_{ij}(\rho)) \to 0 \ \text{ as } \ n \to \infty \ \ \forall i, j \in \{1, 2\}\right\}$$

and the total degrees of freedom $D^\star$ as

$$D^\star \overset{\triangle}{=} \max_{\mathscr{D}}(d^{1,1} + d^{1,2} + d^{2,1} + d^{2,2})$$

The MIMO X channel is especially interesting because it generalizes the interference channel to allow an independent message from each transmitter to each receiver. An interesting coding scheme is proposed by Maddah-Ali et al. (MMK) [15] for the two-user MIMO X channel with three antennas at all nodes. Just as the MIMO X channel combines elements of the MIMO broadcast channel, the MIMO multiple access channel, and the MIMO interference channel into one channel model, the MMK scheme naturally combines dirty thesis coding, successive decoding, and zero-forcing elements into an elegant coding scheme tailored for the MIMO X channel. The results of [38] establish that with three antennas at all nodes the maximum multiplexing gain for each of the MIMO IC, MAC, and BC channels contained within the X channel is 3. However, for the MIMO X channel with three antennas at all nodes, the MMK scheme is able to achieve 4 degrees of freedom. The MMK scheme also extends easily to achieve $4M$ degrees of freedom on the MIMO X channel with $3M$ antennas at each node. Thus, the results of [38] show that the degrees of freedom on the MIMO X channel surpass what is achievable on the interference, multiple access, and broadcast components individually.

Several interesting questions arise for the MIMO X channel. In [20] it is shown that the maximum multiplexing gain for the MIMO X channel, in particular, cannot be achieved by the MMK scheme, but in several setups the MMK scheme is always optimal. They noted that neither dirty thesis coding nor successive decoding has ever been found to be necessary to achieve the full degrees of freedom on any multiuser MIMO channel with perfect channel knowledge. Zero-forcing suffices to achieve all degrees of freedom on the MIMO MAC, BC, and interference channels. They also investigated that the factor of $\frac{4M}{3}$ suggested by the results of Maddah-Ali et al. [15] is found to be optimal; it would lead to non-integer values for the degrees of freedom when M is not an integer multiple of 3. This was of fundamental interest because there were no known results for the optimality of non-integer degrees of freedom for any non-degenerate wireless network with perfect channel knowledge since this research was done. [2] Finally, while the interference channel does not seem to benefit from cooperation through noisy channels between transmitters and receivers, it is not known if shared messages (in the manner of cognitive radio [44]) can improve the degrees of freedom on the MIMO X and interference channels. They used a method called symbol extension over time in order to characterize the total DoF of MIMO X channels, and surprisingly their results were the same as those of the MMK scheme.

What made their results non-practical was that they assumed channel changes over time and that all the transmitters have knowledge all channel states non-causally. Surprisingly, none of above-mentioned methods (neither MMK scheme nor symbol extension over time) could characterize the total DoF of the SISO X channel. Recently a brilliant idea has been put forward to characterize the DoF of the SISO X channel [21], and the $\frac{4}{3}$ of DoF is achieved. In this thesis the results of this method, *Real Alignment* are analyzed. I will describe this method in much more detail. (see section 2.3.4)

2.2.1 Rational Dimension

Proposed in [15], the first example of interference alignment is done in Euclidean spaces. Briefly, the n-dimensional Euclidean space ($n \geq 2$) available at a receiver is partitioned into two subspaces. A subspace is dedicated to interference and all interfering users are forced to respect this constraint. The major technique is to reduce the dimension of this subspace so that the available dimension in the signal subspace allows a higher data rate for the intended users. Alignment using structural codes is also considered by several researchers [17, 35]. Structural interference alignment is used to make the interference caused by users less severe by reducing the number of possible codewords at receivers. Even though useable in one-dimensional spaces, this technique does not allow transmission of different data streams, as there is only one dimension available for transmission.

Maddah-Ali et al. [15] show that there exist available dimensions (called rational dimensions) in one-dimensional spaces, which open new ways of transmitting several data streams

[2] Degrees of freedom with channel uncertainty have been explored in [39, 40, 41, 42, 43].

from a transmitter and interference alignment at the receiver. A coding scheme that provides sufficient tools to incorporate the rational dimensions in transmission is proposed. This coding scheme relies on the fact that irrational numbers can play the role of directions in Euclidean spaces and data can be sent by using rational numbers. This fact is proved by using the results of Hurwitz, Khintchine, and Groshev [45] obtained in the field of Diophantine approximation. In the encoding part, two types of constellation are used to modulate data streams. Type I or single-layer constellation refers to the case where all integer points in an interval are chosen as constellation points. Despite its simplicity, it was shown that the single-layer constellation is capable of achieving the total DoF of several channels. Type II or multi-layer constellation refers to the case where a subset of integer points in an interval is chosen as constellation points. Because it is capable of achieving the total DoF of some channels, this constellation is more useful when all channel gains are rational.

2.3 Coding Scheme

In this section, a coding scheme for data transmission in a shared medium is described in detail. It is assumed that the channel is real, additive, and time invariant. Additive white Gaussian noise with variance σ^2 is added to the received signals at all receivers. Moreover, transmitters are subject to the power constraint P. The SNR is defined as SNR $= \frac{P}{\sigma^2}$.

2.3.1 Transmitting using Rational Dimensions

A transmitter limits its input symbols to a finite set that is called the transmit constellation. Even though it has access to the continuum of real numbers, restriction to a finite set has the benefit of easy and feasible interference management. Having a set of finite points as input symbols, however, does not rule out transmission of multiple data streams from a single transmitter. In what follows, it is shown how a finite set of points can accommodate different data streams.

Let us first explain the encoding of a single data stream. For the sake of simplicity the desired message for the first receiver is noted as $\boldsymbol{u}^j$=$(u_1^j, u_2^j, ..., u_M^j)^t$, and the desired message at the second receiver is noted as $\boldsymbol{v}^j = (v_1^j, v_2^j, ..., v_M^j)^t$.

The transmitter i selects two constellations, $\mathcal{U}^i$ and $\mathcal{V}^i$, to send the data stream i to both receivers. The constellation points are chosen from integer points, i.e., $\mathcal{U}^i \subset \mathbb{Z}^M$ and $\mathcal{V}^i \subset \mathbb{Z}^M$. It is assumed that $\mathcal{U}^i$ and $\mathcal{V}^i$ are bounded sets. Hence, there is a constant Q such that $\mathcal{U}^i \subset [-Q, Q]$ and $\mathcal{V}^i \subset [-Q, Q]$ intervals. The maximum cardinality of $\mathcal{U}^i$ and $\mathcal{V}^i$, which limits the rate of data stream i, is denoted by $|\mathcal{X}^i| = \max\{|\mathcal{U}^i|, |\mathcal{V}^i|\}$. This design corresponds to the case where all integers between $-Q$ and Q are selected, which is a simple choice yet capable of achieving the total DoF of several channels.

Having formed the constellation, the transmitter constructs two random codebooks for data stream j with rates $R^{j,1}$ and $R^{j,2}$ to be received by the first and the second receivers, respectively. This can be accomplished by choosing a probability distribution on the input alphabets. The uniform distribution is the first candidate, and it is selected here for the sake of brevity. On the other hand, since the input constellation is symmetrical by assumption, the expectation of the uniform distribution is zero, and the transmit signal has no DC component. The power consumed by the data stream i can be bounded as Q^2. Even though it is far from being tight in general, using this bound does not decrease the performance of the system as far as the DoF is concerned.

When a transmitter is to send one data stream to each of two receivers, first it uses the above procedure to construct two data streams Then it combines them using a linear combination of all data streams. The transmit signal at the lth antenna of transmitter j can be represented by

$$x_l^j = a_l^j u_l^j + b_l^j v_l^j$$

where u_l^j carries the part of the information for data stream j that is desired at the first receiver and is being transmitted by the lth antenna of transmitter j. Accordingly, v_l^j is defined as the part of the information for data stream j that is desired at the second receiver and is being transmitted by the lth antenna of transmitter j. The a_l^j and b_l^j are constant real numbers that function as separators splitting the two part of data stream j from the transmit signal.

Real numbers a_l^j and b_l^j are rationally independent, i.e., the equation $a_l^j x_1 + b_l^j x_2 = 0$ has no rational solutions for each $j \in \{1, 2, ..., K\}$ and $l \in \{1, 2, ..., M\}$. This independence is because a unique map from constellation points to the message sets is required. Reliance on this independence means that any real number x_l^j belonging to the set of constellation points is uniquely decomposable as $x_l^j = a_l^j u_l^j + b_l^j v_l^j$. Observe that if there is another possible decomposition $x_l^j = \hat{a}_l^j u_l^j + \hat{b}_l^j v_l^j$, then it forces $\hat{a}_l^j$ and $\hat{b}_l^j$ to be rationally dependent.

With the above method each transmitter forms its transmitted data stream $\hat{\boldsymbol{x}}^j = (a_l^j u_l^j + b_l^j v_l^j)$ for $l = 1, 2, ..., M$. To adjust the power, the transmitter multiplies the signal by a constant A, i.e., the transmit signal is $\boldsymbol{x}^j = A\ \hat{\boldsymbol{x}}^j$.

2.3.2 Recovering the Mixed Signal in Rational Dimensions

After rearrangement of the interference part of the signal, the received signal can be represented as

$$y = \hat{g}^0 u^0 + \hat{g}^1 I^1 + \ldots + \hat{g}^m I^m + z \tag{2.4}$$

where $\hat{g}^0 = g^0$ to unify the notation. In what follows, the decoding scheme used to decode u^0 from y is explained. It is worth noting that if the receiver is interested in more than one data stream, then it performs the same decoding procedure for each data stream.

At the receiver, the received signal is first passed through a hard decoder. The hard decoder looks at the received constellation $\hat{\mathcal{U}} = g^0 \mathcal{U}^0 + \hat{g}^1 \mathcal{I}^1 + \ldots + \hat{g}^m \mathcal{I}^m$ and maps the received signal to the nearest point in the constellation. This changes the continuous channel to a discrete one in which the input symbols are from the transmit constellation $\mathcal{U}^0$ and the output symbols are from the received constellation.

Remark 1 *$\mathcal{I}^j$ is the constellation due to single or multiple data streams. Since it is assumed that in the latter case there is a linear combination of multiple data streams with integer coefficients, it can be concluded that $\mathcal{I}^j \subset \mathbb{Z}$ for $j \in \{1, 2, \ldots, m\}$.*

To bound the performance of the decoder, it is assumed that the received constellation has the property that there is a many-to-one map from $\hat{\mathcal{U}}$ to $\mathcal{U}^0$. This in fact implies that if there is no additive noise in the channel then the receiver can decode the data stream with zero error probability. This property is called property Γ. It is assumed that this property holds for all received constellations. To satisfy this requirement at all receivers, usually a careful transmit constellation design is needed at all transmitters.

Let $d_{\min}$ denote the minimum distance in the received constellation. Having property Γ, the receiver passes the output of the hard decoder through the many-to-one map from $\hat{\mathcal{U}}$ to $\mathcal{U}^0$. The output is called $\hat{u}^0$. Now, a joint-typical decoder can be used to decode the data stream from a block of $\hat{u}^0$. To calculate the achievable rate of this scheme, the error probability of transmitting a symbol from $\mathcal{U}^0$ and receiving another symbol, i.e., $P_e = Pr\{\hat{U}^0 \neq U^0\}$, is bounded as

$$P_e \quad \leq Q\left(\tfrac{d_{\min}}{2\sigma}\right) \leq \exp\left(-\tfrac{d_{\min}^2}{8\sigma^2}\right) \tag{2.5}$$

Definition 1 (Noise Removal) *A receiver can completely remove the noise if the minimum distance between the received constellation points is greater than $\sqrt{N}$, where N is the noise variance [24].*

Now P_e can be used to lower bound the rate achievable for the data stream. Etkin and Ordentlich [36] used Fano's inequality to obtain a lower bound on the achievable rate, which is tight in high Signal-to-Noise Ratio (SNR) regimes. Following similar steps, one can obtain

$$\begin{aligned} R \quad &= I(\hat{u}^0, u^0) \\ &= H(u^0) - H(u^0|\hat{u}^0) \\ &\overset{a}{\geq} H(u^0) - 1 - P_e \log|\mathcal{U}^0| \\ &\overset{b}{\geq} \log|\mathcal{U}^0| - 1 - P_e \log|\mathcal{U}^0| \end{aligned} \tag{2.6}$$

where (a) follows from Fano's inequality and (b) follows from the fact that u^0 has uniform distribution. To have a multiplexing gain of at least d, $|\mathcal{U}^0|$ needs to scale as SNR^d. Moreover, if P_e scales as $\exp(\mathrm{SNR}^{-\epsilon})$ for an $\epsilon > 0$, then it can be shown that $\frac{R}{\log \mathrm{SNR}}$ approaches d at high SNR regimes.

It is noteworthy that after interference alignment the interference term no longer has uniform distribution. However, the lower bound on the achievable rate given in (2.6) is independent of the probability distributions of the interference terms. It is possible to obtain better performance by exploiting the distribution of the interference.

2.3.3 Performance Analysis: Khintchine–Groshev Theorem

The decoding scheme proposed in the previous section is used to decode the data stream u_0 from the received signal in (2.4). To satisfy property Γ, it is assumed that $\{\hat{g}_0, \hat{g}_1, \ldots, \hat{g}_m\}$ are independent over rational numbers. Owing to this independence, any point in the received constellation has a unique representation in the bases $\{\hat{g}_0, \hat{g}_1, \ldots, \hat{g}_m\}$, and therefore property Γ holds in this case.

Remark 2 *In a random environment it is easy to show that the set of* $\{\hat{g}_0, \hat{g}_1, \ldots, \hat{g}_m\}$ *, being dependent, has measure zero (with respect to Lebesgue measure). Hence, in this section it is assumed that property* Γ *holds unless otherwise stated.*

To use the lower bound on the data rate given in (2.6), one needs to calculate the minimum distance between points in the received constellation. Let us assume each stream in (2.4) is bounded (as is the case, since transmit constellations are bounded by the assumption). In particular, $\mathcal{U}_0 = [-Q_0, Q_0]$ and $\mathcal{I}_j = [-Q_j, Q_j]$ for all $j \in \{1, 2, \ldots, m\}$. Since points in the received constellation are irregular, finding $d_{\min}$ is not easy in general. Thanks to the theorems of Khintchine and Groshev [46], however, it is possible to lower bound the minimum distance. As is shown in [24], using this lower bound at high SNR regimes is asymptotically optimal. Some background needed for stating the theorem of Khintchine and Groshev will now be provided.

The field of Diophantine approximation in number theory deals with approximation of real numbers with rational numbers. The reader is referred to [47, 48] and the references therein. The Khintchine theorem is one of the cornerstones in this field. It gives a criteria for a given function $\psi : \mathbb{N} \to \mathbb{R}_+$ and real number α such that $|p + \alpha q| < \psi(|q|)$ has either infinitely many solutions or at most finitely many solutions for $(p, q) \in \mathbb{Z}^2$. Let $\mathcal{A}(\psi)$ denote the set of real numbers such that $|p + \alpha q| < \psi(|q|)$ has infinitely many solutions in integers. The theorem has two parts. The first part is the convergent part and states that if $\psi(|q|)$ is convergent, i.e.,

$$\sum_{q=1}^{\infty} \psi(q) < \infty$$

then $\mathcal{A}(\psi)$ has measure zero with respect to Lebesque measure. This part can be rephrased in a more convenient way as follows. For almost all real numbers, $|p + \alpha q| > \psi(|q|)$ holds for all $(p,q) \in \mathbb{Z}^2$ except for a finitely many of them. Since the number of integers violating the inequality is finite, one can find a constant κ such that

$$|p + \alpha q| > \kappa\psi(|q|)$$

holds for all integers p and q almost surely. The divergent part of the theorem states that $\mathcal{A}(\psi)$ has the full measure, i.e., the set $\mathbb{R} - \mathcal{A}(\psi)$ has measure zero, provided ψ is decreasing and $\psi(|q|)$ is divergent, i.e.,

$$\sum_{q=1}^{\infty} \psi(q) = \infty.$$

There is an extension to Khintchine's theorem that looks at the approximation of linear forms. Let $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_m)$ and $\boldsymbol{q}= (q_1, q_2, \ldots, q_m)$ denote an m-tuple in $\mathbb{R}^m$ and $\mathbb{Z}^m$, respectively. Let $\mathcal{A}_m(\psi)$ denote the set of m-tuple real numbers $\boldsymbol{\alpha}$ such that

$$|p + \alpha_1 q_1 + \alpha_2 q_2 + \ldots + \alpha_m q_m| < \psi(|\boldsymbol{q}|_\infty) \tag{2.7}$$

has infinitely many solutions for $p \in \mathbb{Z}$ and $\boldsymbol{q} \in \mathbb{Z}^m$. The $|\boldsymbol{q}|_\infty$ is the supreme norm of defined as $\max_i |q_i|$. The following theorem gives the Lebesque measure of the set $\mathcal{A}_m(\psi)$.

Theorem 1 (Khintchine–Groshev) *Let $\psi : \mathbb{N} \to \mathbb{R}^+$. Then the set $\mathcal{A}_m(\psi)$ has measure zero provided*

$$\sum_{q=1}^{\infty} q^{m-1}\psi(q) < \infty, \tag{2.8}$$

and has the full measure if

$$\sum_{q=1}^{\infty} q^{m-1}\psi(q) = \infty \tag{2.9}$$

and ψ is monotonic.

In real interference alignment, the convergent part of the theorem is concerned, which can be proved using the Borel–Cantelli lemma. Moreover, given an arbitrary $\epsilon > 0$, the function $\psi(q) = \frac{1}{q^{m+\epsilon}}$ satisfies (2.8). In fact, the convergent part of the theorem used in this section can be stated as follows. For almost all m-tuple real numbers there exists a constant κ such that

$$|p + \alpha_1 q_1 + \alpha_2 q_2 + \ldots + \alpha_m q_m| > \frac{\kappa}{(\max_i |q_i|)^{m+\epsilon}} \tag{2.10}$$

holds for all $p \in \mathbb{Z}$ and $\mathbf{q} \in \mathbb{Z}^m$.

The Khintchine–Groshev theorem can be used to bound the minimum distance of points in the received constellation. In fact, a point in the received constellation has a linear form, i.e., $u_r = g_0 u_0 + \hat{g}_1 I_1 + \ldots + \hat{g}_m I_m$. Dividing by g_0 and using (2.10), one can conclude that

$$d_{\min} > \frac{\kappa g_0}{(\max_{i\in\{1,\ldots,m\}} Q_i)^{m+\epsilon}} \tag{2.11}$$

The probability of error in hard decoding, see (2.5), can be bounded as

$$P_e < \exp\left(-\frac{(\kappa g_0)^2}{8\sigma^2(\max_{i\in\{1,\ldots,m\}} Q_i)^{2m+2\epsilon}}\right) \tag{2.12}$$

Let us assume Q_i for $i \in \{0, 1, \ldots, m\}$ is $\lfloor \gamma_i P^{\frac{1-\epsilon}{2(m+1+\epsilon)}} \rfloor$ where γ_i is a constant. Moreover, ϵ is the constant appearing in (2.10). Let us also assume that $g_0 = \gamma P^{\frac{m+2\epsilon}{2(m+1+\epsilon)}}$. It is worth mentioning that in this thesis it is assumed that each data stream carries the same rate in the asymptotic case of high SNR, i.e., they have the same multiplexing gain. However, in more general cases one may consider different multiplexing gains for different data streams. Substituting in (2.12) yields

$$P_e < \exp\left(-\delta P^{\epsilon}\right) \tag{2.13}$$

where δ is a constant and a function of γ, κ, σ, and γ_i. The lower bound obtained in (2.6) for the achievable rate becomes

$$\begin{aligned} R_0 \quad &> (1 - P_e)\log|\mathcal{U}_0| - 1 \\ &\overset{a}{=} (1 - \exp(-\delta P^{\epsilon}))\log(2\lfloor \gamma_i P^{\frac{1-\epsilon}{2(m+1+\epsilon)}} \rfloor) - 1 \\ &> \frac{(1-\epsilon)(1 - \exp(-\delta P^{\epsilon}))}{2(m+1+\epsilon)}(\log(P) + \vartheta) - 1 \end{aligned} \tag{2.14}$$

where (a) follows from the fact that $|\mathcal{U}_0| = 2Q_0$ and ϑ is a constant. The multiplexing gain of the data stream u_0 can be computed using (2.14) as follows

$$\begin{aligned} R_0 \quad &= \lim_{P\to\infty} \frac{R_0}{0.5\log(P)} \\ &> \frac{1-\epsilon}{m+1+\epsilon}. \end{aligned} \tag{2.15}$$

Since ϵ can be made arbitrarily small, it can be concluded that $d = \frac{1}{m+1}$ is indeed achievable. In [24], the following theorem is proved; this result and its required conditions are summarized.

Theorem 2 *A receiver can reliably decode the data stream u_0 with multiplexing gain $\frac{1}{m+1}$ from the received signal $y = g_0 u_0 + \hat{g}_1 I_1 + \ldots + \hat{g}_m I_m + z$ if the following regularity conditions are satisfied:*

1. $g_0 = \gamma P^{\frac{m+2\epsilon}{2(m+1+\epsilon)}}$ *where γ is a constant.*

2. $u_0 \in [-Q_0, Q_0]$ *where* $Q_0 = \lfloor \gamma_0 P^{\frac{1-\epsilon}{2(m+1+\epsilon)}} \rfloor$ *and γ_0 is a constant. Moreover, the uniform distribution is used to construct the random codebook.*

3. *For $i \in \{1, 2, \ldots, m\}$, $I_i \in [-Q_i, Q_i]$ where $Q_i = \lfloor \gamma_i P^{\frac{1-\epsilon}{2(m+1+\epsilon)}} \rfloor$ and γ_i is a constant.*

4. *$\hat{g}_i$s for $i \in \{0, 1, \ldots, M\}$ are independent over rational numbers.*

5. $\{\frac{\hat{g}_1}{g_0}, \frac{\hat{g}_2}{g_0}, \ldots, \frac{\hat{g}_m}{g_0}\}$ *is among m-tuples that satisfy (2.10).*

Moreover, the last two conditions almost surely hold.

2.3.4 $K \times M$ SISO X Channel

System Model

The $K \times M$ SISO X channel models a network in which K transmitters wish to communicate with M receivers. Unlike the interference channel, each transmitter has a message for each receiver. In other words, transmitter i for all $i \in \{1, 2, \ldots, K\}$ wishes to transmit an independent message to receiver j for all $j \in \{1, 2, \ldots, M\}$. The message transmitted by transmitter i and intended for receiver j is denoted by m_{ij}. The channel's input–output relationship can be stated as follows, see Figure 2.2 (User i for $i \in \{1, 2, \ldots, K\}$ wishes to transmit an independent message m_{ij} to receiver j for all $j \in \{1, 2, \ldots, M\}$.),

$$\begin{aligned} y^1 &= h^{1,1}x^1 + h^{2,1}x^2 + \ldots + h^{K,1}x^K + z^1, \\ y^2 &= h^{1,2}x^1 + h^{2,2}x^2 + \ldots + h^{K,2}x^K + z^2, \\ \vdots &= \ \vdots \qquad \vdots \qquad \ddots \qquad \vdots \\ y^M &= h^{1,M}x^1 + h^{2,M}x^2 + \ldots + h^{K,M}x^K + z^M, \end{aligned} \tag{2.16}$$

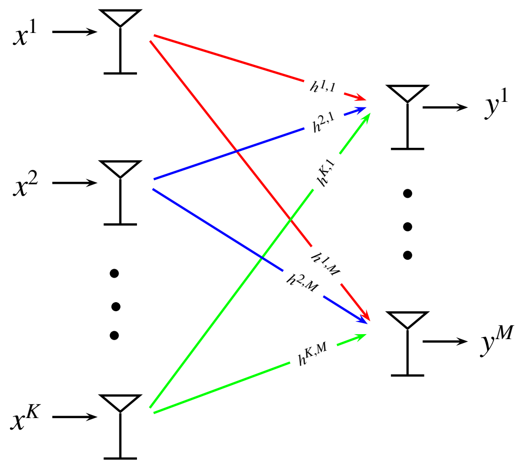

Figure 2.2: The $K \times M$ SISO X Channel

where x^i and y^i are input and output symbols of user i for $i \in \{1, 2, \ldots, K\}$, respectively. z^i is AWGN with unit variance for $i \in \{1, 2, \ldots, K\}$. Transmitters are subject to the power constraint P. The $h^{i,j}$ represents the channel gain between transmitter i and receiver j. It is assumed that all channel gains are real and time invariant.

Let C denote the capacity region of this channel. The DoF region associated with the channel can be defined as the shape of the region in high SNR regimes scaled by log SNR. Let us denote the DoF region by $\mathscr{D}$. Of primarily interest is the main facet of the DoF region, defined as

$$D = \lim_{\text{SNR}\to\infty} \max_{\mathbf{R}\in C} \frac{\sum_{i=1}^{K} \sum_{j=1}^{M} R^{i,j}}{\log \text{SNR}} \tag{2.17}$$

where $R^{i,j}$ is an achievable rate for the message m_{ij} and $\mathbf{R}$ is the set of all achievable rates. The DoF achievable by the message m_{ij} is denoted by $d^{i,j}$.

The Total DoF of $\frac{KM}{K+M-1}$ is Achievable

An upper bound on the DoF of this channel is obtained in [23]. The upper bound states that the total DoF of the channel is less than or equal to $\frac{KM}{K+M-1}$, which means each message can at most achieve $\frac{1}{K+M-1}$ of DoF. It will be shown that this DoF is achievable. To this end, transmitter i for all $i \in \{1, 2, \ldots, K\}$ transmits M signals along M directions as follows:

$$x^i = \sum_{j=1}^{M} h^{i,j} x^{i,j} \tag{2.18}$$

where $x^{i,j}$ is the signal carrying the message m_{ij}. Let us focus on the signals intended for receiver 1, i.e., $x^{1,1}, x^{2,1}, \ldots, x^{K,1}$. The received signals due to these transmit signals can be written as

$$\begin{aligned}
\tilde{y}^1 &= (h^{1,1})^2 x^{1,1} + (h^{2,1})^2 x^{2,1} + \ldots + (h^{K,1})^2 x^{K,1} \\
I^{1,2} &= (h^{1,2}h^{1,1})x^{1,1} + (h^{2,2}h^{2,1})x^{2,1} + \ldots + (h^{K,2}h^{K,1})x^{K,1} \\
\vdots &= \vdots \qquad \vdots \qquad \ddots \qquad \vdots \\
I^{1,M} &= (h^{1,M}h^{1,1})x^{1,1} + (h^{2,M}h^{2,1})x^{2,1} + \ldots + (h^{1,M}h^{K,1})x^{K,1}.
\end{aligned} \tag{2.19}$$

Since $x^{1,1}, x^{2,1}, \ldots, x^{K,1}$ are not intended for receiver j for all $j \in \{2, 3, \ldots, M\}$, $I^{1,j}$ is a part of interference at receiver j. It is claimed that all interfering signals $x^{1,1}, x^{2,1}, \ldots, x^{K,1}$ can be aligned at all receivers $j \in \{2, 3, \ldots, M\}$. Let $\mathbf{H}_1$ denote the set of all coefficients appearing in $I^{1,2}, I^{1,3}, \ldots, I^{1,M}$, i.e., $\mathbf{H}_1 = \{(h^{1,2}h^{1,1}), (h^{2,2}h^{2,1}), \ldots, (h^{2,M}h^{2,1}), (h^{1,M}h^{K,1})\}$. $\mathbf{H}_1$ has $(M-1)K$ members. The set of all monomials with variables in $\mathbf{H}_1$ is denoted by $\mathcal{G}(\mathbf{H}_1)$. Let $\mathcal{T}_1$ denote a subset of $\mathcal{G}(\mathbf{H}_1)$ consisting of monomials represented by

$$\mathcal{T}_1 = \prod_{i=1}^{K} \prod_{j=1}^{M} (h^{i,j}h^{i,1})^{s_{i,j}} \tag{2.20}$$

where

$$\begin{cases} s_{i,1} = 0 & \forall\, i \in \{1, 2, \ldots, K\} \\ 0 \leq s_{i,j} \leq n & \text{Otherwise} \end{cases}$$

Clearly, $\mathcal{T}_1$ has $(n+1)^{(M-1)K}$ members.

The message $m_{i,1}$ for $i \in \{1, 2, \ldots, K\}$ is transmitted along directions in $\mathcal{T}_{i,1}$ where $\mathcal{T}_{i,1} \subset \mathcal{T}_1$. A direction T in $\mathcal{T}_{i,1}$ can be represented as

$$T = \prod_{l=1}^{K} \prod_{j=1}^{M} (h^{l,j}h^{l,1})^{s_{l,j}} \tag{2.21}$$

where

$$\begin{cases} s_{l,1} = 0 & \forall\, l \in \{1, 2, \ldots, K\} \\ 0 \leq s_{i,j} \leq n-1 & \forall\, j \in \{1, 2, \ldots, M\}\ \&\ j \neq 1 \\ 0 \leq s_{l,j} \leq n & \text{Otherwise} \end{cases}$$

It is easy to show that the cardinality of $\mathcal{T}_{i,1}$ is $n^{M-1}(n+1)^{(M-1)(K-1)}$. The received directions due to $x^{i,1}$ at all receivers belong to $\mathcal{T}_1$. In fact, $x^{i,1}$ arrives at receiver j multiplied by $(h^{i,j}h^{i,1})$, and since the power of $(h^{i,j}h^{i,1})$ in all directions in $x^{i,1}$ is less than n, it is concluded that the received directions are all in $\mathcal{T}_1$. Therefore, all transmit signals are aligned, and the total number of directions in $I^{1,j}$ for all $j \in \{2, 3, \ldots, M\}$ is $(n+1)^{(M-1)K}$.

A similar argument can be applied for signals intended for receiver j for all $j \in \{2, 3, \ldots, M\}$. Therefore, the received signals can be represented as

$$\begin{aligned}
y^1 &= \tilde{y}^1 + I^{2,1} + I^{3,1} + \ldots + I^{M,1} + z^1, \\
y^2 &= \tilde{y}^2 + I^{1,2} + I^{3,2} + \ldots + I^{M,2} + z^2, \\
\vdots &= \vdots \qquad \vdots \qquad \ddots \qquad \vdots \\
y^M &= \tilde{y}^M + I^{1,M} + I^{2,M} + \ldots + I^{M,(M-1)} + z^M,
\end{aligned} \tag{2.22}$$

where $I^{i,j}$ is the part of interference caused by all messages intended for receiver i at receiver j. Because of symmetry, only the received directions at receiver 1 are considered. At receiver 1, there are $(M-1)$ interfering signals each of which consists of at most $(n+1)^{(M-1)K}$ directions. Therefore, the total number of interfering directions is $L'_1 = (M-1)(n+1)^{(M-1)K}$. On the other hand, $\tilde{y}^1$ consists of $Kn^{M-1}(n+1)^{(M-1)(K-1)}$ directions. This is because $\tilde{y}^1 = (h^{1,1})^2x^{1,1} + (h^{2,1})^2x^{2,1} + \ldots + (h^{K,1})^2x^{K,1}$ and $x^{i,1}$ for all $i \in \{1, 2, \ldots, K\}$ consists of $n^{M-1}(n+1)^{(M-1)(K-1)}$ directions. Therefore, the total number of received directions is

$$L = (M-1)(n+1)^{(M-1)K} + Kn^{M-1}(n+1)^{(M-1)(K-1)}.$$

The use of Theorem 2, allows the conclusion that

$$D \geq \frac{KMn^{M-1}(n+1)^{(M-1)(K-1)}}{Kn^{M-1}(n+1)^{(M-1)(K-1)} + (M-1)(n+1)^{(M-1)K} + 1} \tag{2.23}$$

is achievable for the X channel. Rearranging gives

$$D \geq \frac{KM}{K + (M-1)\left(\frac{n+1}{n}\right)^{M-1} + \frac{1}{n^{M-1}(n+1)^{(M-1)(K-1)}}} \tag{2.24}$$

Since (2.24) holds for all n,

$$D = \frac{KM}{K+M-1} \tag{2.25}$$

which is the desired result. In a special case, $M = K = 2$ and the total DoF is $\frac{4}{3}$. On the other hand, for the general case of $M = K$ the conclusion is that the total DoF of $\frac{K^2}{2K-1}$, which is always greater than $\frac{K}{2}$, and it can be observed that as the number of receivers increases the total DoF approaches $\frac{K}{2}$, which means that X channel acts like GIC in if the merit is DoF.

2.4 General Case Scenario

In the general and more practical situation there are always more than two transmitters and receivers with different number of antennas at each node. All of the aforementioned techniques fall short even in MIMO setups with an equal number of antennas at all nodes. In the upcoming chapters a novel scheme, called layered interference alignment, will be introduced. This was first proposed by me in [27] and could achieve the total DoF of a class of MIMO X channels. One of the main advantages of this novel technique in comparison with MMK and symbol extension methods is its applicability for both time invariant and complex channel realizations. In the last chapter, I will outline the clues to solve much more general setups. The layered interference alignment technique is a powerful scheme, which will spur further research on sophisticated interference management in wireless networks.

Chapter 3

Mathematical Analysis Tools

In this chapter several tools from the field of Diophantine approximation are discussed. Using the integers as the desired transmit constellations and designing the transmit signals based on rational dimensions spanned by the channel gains makes the need for approximation of integers inevitable. The tools that will be discussed here bring the possibility of decoder analysis at high SNR regimes using a simple hard decoder. This also provides required tools for the proof of essential mathematical theorems of simultaneous/joint decoding in the layered interference alignment.

3.1 Integer Number Estimation

Diophantine approximation is a branch of number theory that can loosely be described as a quantitative analysis of the property that every real number can be approximated by a rational number arbitrarily closely. The theory dates back to the ancient Greeks and Chinese who used good rational approximations to the number pi (3.14159...) in order to accurately predict the positions of planets and stars. The metric theory of Diophantine approximation is the study of the approximation properties of real numbers by rationals from a measure theoretic (probabilistic) point of view. The central theme is to determine whether a given approximation property holds everywhere except on an exceptional set of measure zero. Usually the approximation property is used to look for those real numbers that fall into a given interval of length determined by the denominators of the rational approximates infinitely often. Such numbers are called *well-approximable*, and the set can be expressed as a lim sup set.

This branch of number theory takes its roots from Dirichlet's famous theorem, which is based on the box counting argument. Throughout this chapter classical results from Diophantine approximation, which can be found in [25, 49, 50], are stated.

Theorem (Dirichlet) 1 *Given $\alpha \in \mathbb{R}$, $N \in \mathbb{N}$, there exist integers p, q with $1 \leq q \leq N$ such that*

$$\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{q(N+1)}$$

The proof of this is a direct consequence of the pigeon hole principle — given N objects and M boxes with $N > M \geq 1$, then there exists a box with at least two objects. The following global statement concerning the rate of rational approximation to any given irrational number can be deduced from Dirichlet's theorem.

Corollary 1 *For any irrational $\alpha \in \mathbb{R}$, there exist infinitely many integers p and $q > 0$ such that*

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{q^2} \tag{3.1}$$

The corollary simply states that any irrational number can be approximated by rationals with the rate of one over the denominator squared. This greatly improves the trivial rate of one over the denominator, which makes use of the simple fact that for any α and fixed denominator q there exists an integer p such that $|\alpha - p/q| \leq 1/2q$.

A natural question now arises. Is it possible to do better? That is to say, can we replace the right-hand side of (3.1) with a quantity tending to zero faster as the denominator size increases? The following result completely answers this question.

Theorem (Hurwitz 1891) 1 *For any irrational real number α there exist infinitely many integers p and $q > 0$ such that*

$$\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{\sqrt{5}q^2} \tag{3.2}$$

Furthermore, the constant $\sqrt{5}$ in the above inequality is the best possible.

The previous statement simply means that it is not possible to replace $\sqrt{5}$ in (3.2) by a constant A strictly larger than $\sqrt{5}$.

It follows from the theory of continued fractions that numbers equivalent to the golden ratio $\frac{\sqrt{5}+1}{2}$ have the worst approximation properties amongst the irrationals. In particular, for any integers p and $q > 0$

$$\left|\frac{\sqrt{5}+1}{2} - \frac{p}{q}\right| > \frac{1}{(\sqrt{5}+\epsilon)q^2}$$

where $\epsilon > 0$ is arbitrary. To a certain extent, this can be explained by the fact that the golden ratio has the simplest continued fraction expansion, which is

$$\frac{\sqrt{5}+1}{2} = [1; 1, 1, 1, \ldots]$$

Here, the notation $[a_0; a_1, a_2, \ldots]$ denotes the continued fraction expansion of an irrational α, i.e.,

$$\alpha = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \ldots}}$$

.

The partial quotients a_i are integers with $a_i \geq 1$ for $i \geq 1$. This discussion leads naturally to the notion of badly approximable numbers.

A natural question now arises as to whether it is possible to do better? That is, is it possible to improve the right-hand side of inequality 3.1. It follows from continued fractions that q^{-2} can be replaced by $\frac{1}{2q^2}$. In 1891 Hurwitz completely answered it by proving that the best possible by considering the continued fraction expansion of the numbers that are equivalent to golden ratios $\alpha = (\sqrt{5}+1)/2$. That is, $\alpha = (\sqrt{5}+1)/2$ has worst approximation amongst the irrationals, i.e.,

$$\left|\frac{(\sqrt{5}+1)}{2} - \frac{p}{q}\right| > \frac{1}{(\sqrt{5}+\epsilon)q^2}$$

In fact, the above holds true for any numbers equivalent to the golden mean, i.e., numbers of the form

$$y = \frac{ax+b}{cx+d}$$

where $a, b, c, d \in \mathbb{Z}$ and

$$ad - bc = 1$$

where $\epsilon > 0$ is arbitrary. This discussion leads naturally to the notion of *badly approximable numbers*.

A real number α is said to be badly approximable if there exists a constant $c = c(\alpha) > 0$ such that

$$\left|\alpha - \frac{p}{q}\right| > \frac{c}{q^2}$$

for all integers p and $q > 0$. Thus, a number is badly approximable if the rate of approximation given by 3.1 cannot be improved beyond a constant. Let **Bad** denote the set of badly approximable numbers. Clearly **Bad** is non-empty — the golden ratio is in **Bad**. The set **Bad** of badly approximable numbers can be completely described in terms of the theory of continued fractions. A classical result states that α is badly approximable if and only if the partial quotients in its continued fraction expansion are bounded, i.e., there exists a positive constant $K(\alpha)$ such that

$$\alpha \in \mathbf{Bad} \qquad \Longleftrightarrow \qquad \exists\ K(\alpha) > 0\ :\ |a_i| \leq K(\alpha) \quad \forall\ i\ \geq\ 1$$

All the quadratic irrationals ($\sqrt{2},\ \sqrt{3},\ \sqrt{5}, \cdots$) have bounded partial quotients and are therefore badly approximable.

Bad is a small set in terms of measure, as its Lebesgue measure is zero (see Section 3.2), but it has maximal dimension in $\mathbb{R}$.

To reiterate the main point of the above discussion: the exponent 2 in (3.1) is the best possible in the sense that if it is replaced by $\tau > 2$ then (3.1) is no longer valid for all irrationals. This naturally brings us to the notion of τ -*approximable* numbers — numbers for which the exponent 2 in (3.1) can be improved. Let τ be a positive real number. A real number α is said to be τ-approximable if

$$\left|\alpha - \frac{p}{q}\right| \leq q^{-\tau} \qquad \text{for infinitely many} \qquad (p, q) \in \mathbb{Z} \times \mathbb{N}$$

The set of τ-approximable numbers will be denoted by $W(\tau)$. In view of Dirichlet's theorem, we trivially have that

$$W(\tau) = \mathbb{R} \qquad \text{for any} \qquad \tau \leq 2$$

On the other hand there exist very well approximable numbers that can be approximated by rationals to within a rate of $q^{-\tau}$ for an arbitrary large value of τ. Such numbers are called

Liouville numbers. Formally, an irrational α is said to be a Liouville number if

$$\alpha \in \bigcap_{\tau>0} W(\tau)$$

For example, a consequence of Liouville's theorem [25] is that the number

$$10^{-1!} + 10^{-2!} + 10^{-3!} + \dots$$

is a Liouville number. The set of Liouville numbers will be denoted by $\mathcal{L}$. Clearly, $\mathcal{L} \subset W(2+\epsilon)$ for arbitrary $\epsilon > 0$, and it is well known that both badly approximable and Liouville numbers are quite rare. Indeed,

$$|\mathbf{Bad}| \;=\; |\mathcal{L}| =\; 0,$$

where $|X|$ is the Lebesgue measure of the set X. In other words, a randomly chosen real number lies outside these sets with probability 1. However, by another, more delicate notion of "size" **Bad** is bigger than $\mathcal{L}$. More precisely, $\dim(\mathbf{Bad}) \;=\; 1 \;>\; 0 \;=\; \dim \mathcal{L}$, where $\dim(X)$ is the Hausdorff dimension of X. See [49] for further details.

3.2 Khintchine's Theorem

We start by generalizing the notion of τ-approximable numbers by introducing general approximating functions.

Throughout the chapter, an "approximating function" means a decreasing function $\psi\colon \mathbb{N} \mapsto \mathbb{R}^+$ such that $\psi(r) \to 0$ as $r \to \infty$. Let $W(\psi)$ denote the set of all numbers that satisfy

$$|x - p/q| \;\leq\; \psi(q) \qquad \text{for infinitely many} \qquad (p, q) \in \mathbb{Z} \times \mathbb{N}$$

Thus, with $\psi : q \mapsto \psi(q) := q^{-\tau}$ the general set $W(\psi)$ is simply the set $W(\tau)$ of τ-approximable numbers.

A straightforward application of the Borel–Cantelli lemma shows that if $\sum_{q=1}^{\infty} q\psi(q) \;<\; \infty$ then

$$|W(\psi)| = 0$$

Figure 3.1: Khintchine 1-dimensional approximation function

Thus, if the function ψ decreases rapidly enough so that the "measure" sum converges, then the set of ψ-approximable numbers is of zero Lebesgue measure.

A natural question now arises. What can be said about the size of the set $W(\psi)$ when the measure sum diverges? The answer to this is given by Khintchine in [51], and it is a key result in metric Diophantine approximation. It provides a beautiful and simple "zero-full" criterion for the size of $W(\psi)$ expressed in terms of Lebesgue measure.

Theorem (Khintchine) 1 *Let ψ be an approximating function. Then,*

$$|W(\psi) \cap \mathbb{I}| = \begin{cases} 0 & \text{if} \quad \sum_{q=1}^{\infty} q\psi(q) < \infty \\ \\ 1 & \text{if} \quad \sum_{q=1}^{\infty} q\psi(q) = \infty \end{cases}$$

where $\mathbb{I} := [-\frac{1}{2}, \frac{1}{2}]$ throughout this chapter unless or otherwise specified. The reason for this choice will be apparent later. For understanding one may consider the unit interval $\mathbb{I} = [0, 1]$, and in this case Figure 3.1 may be considered to understand Khintchine's theorem.

Remark 3 *The fact that the statement is for $W(\psi) \cap \mathbb{I}$, rather than simply $W(\psi)$, is for convenience and is actually stronger. It implies that when the measure sum diverges the set $W(\psi)$ is of full Lebesgue measure, i.e., the complementary set $\mathbb{R} \setminus W(\psi)$ is of zero Lebesgue measure. Naturally this is stronger than the statement that $|W(\psi)| = \infty$.*

As already mentioned, the convergence case follows easily the simple application of the Borel–Cantelli lemma. Thus, the divergence case constitutes the main substance of the theorem, and that is where the assumption ψ is decreasing comes into play. It is worth mentioning that in Khintchine's original statement the stronger hypothesis that $r^2\psi(r)$ is decreasing was assumed.

The fact that this additional hypothesis is unnecessary is due to [52]. However, removing the hypothesis that ψ is decreasing in Khintchine's theorem represents a fundamental open problem in metric number theory. This condition cannot in general be relaxed, as was shown by Duffin and Schaeffer [53] in 1941. They constructed a function $\psi : \mathbb{N} \to \mathbb{R}^+$ (non-monotonic approximating function) for which the sum $\sum_{q=1}^{\infty} q\psi(q)$ diverges but $|W(\psi)| = 0$. Nevertheless, Duffin and Schaeffer produced the following conjecture for an arbitrary ψ. Let

$$W'(\psi) = \left\{x \in \mathbb{I} : \left|x - \frac{p}{q}\right| < \psi(q) \text{ for i.m. } p \in \mathbb{Z}, q \in \mathbb{N} \ \ \& \ \ (p,q) = 1\right\} \tag{3.3}$$

Obviously

$$W'(\psi) \subset W(\psi)$$

The co-primeness condition is imposed in (3.3) in order to relate the rationals to the approximating function uniquely, as it is irrelevant when ψ is monotonic. Let ϕ denote the *Euler function*; then the conjecture is as follows.

Duffin-Schaeffer **Conjecture**: *For any function* $\psi : \mathbb{N} \mapsto \mathbb{R}^+$

$$|W'(\psi)| = 1 \quad if \quad \sum_{q=1}^{\infty} \varphi(q)\psi(q) \ = \ \infty$$

Although various partial results have been established (see [54] for details and references), the full conjecture represents one of the most difficult and profound unsolved problems in metric number theory.

Khintchine's theorem is a very delicate statement. For example, with $\psi_1(q) = \frac{1}{q^2 \log q}$ the measure sum diverges whereas with $\psi_2(q) = \frac{1}{q^2(\log q)^{1+\epsilon}}$ and ($\epsilon > 0$) the sum converges. Thus, an increase in the rate of approximation by a factor of $(\frac{1}{\log q})^{\epsilon}$ switches the measure from full to zero, i.e.,

$$|W(\psi_1) \cap \mathbb{I}| = 1 \qquad \text{and} \qquad |W(\psi_2) \cap \mathbb{I}| = 0$$

Let us return briefly to the set **Bad** of badly approximable numbers. It follows from the definition of **Bad** and $W(\psi)$ that

$$\textbf{Bad} \subset \mathbb{I} \setminus W(\psi_1)$$

The set on the right-hand side is of Lebesgue measure zero, and thus

$$|\mathbf{Bad}| = 0$$

Now, for any decreasing function ψ tending to zero faster than ψ_2, $|W(\psi)| = 0$. In particular, $|W(10)| = 0$ and $|W(100)| = 0$, and Khintchine's theorem fails to distinguish between them. Heuristically, one would expect that the set $W(100)$ is in some sense smaller than the set $W(10)$. But how can this be characterized? One requires another notion of size suited for describing the finer measure theoretic structure of $W(\psi)$ — beyond Lebesgue measure. The appropriate notion of size for this purpose is the notion of generalized Hausdorff measures. A curious reader who wants to go into further details of metric Diophantine approximation is referred to [55]. However, later in this thesis the Hausdorff dimension of badly approximable sets will be calculated. Therefore, a brief introduction of the Hausdorff dimension and measure is now presented.

3.3 Hausdorff Measure and Dimension

The Hausdorff dimension of a set X is a generalization of the standard concept of dimension. To define the Hausdorff dimension of the set X the notion of Hausdorff measure is required. Just as two subsets of $\mathbb{R}^n$ may have equal dimension but differing n-dimensional volume, two subsets of X may have equal Hausdorff dimension but different Hausdorff measure. Furthermore, the Hausdorff measure agrees with the standard Lebesgue definition of measure for Lebesgue measurable sets. Thus, Hausdorff measure is a generalization of Lebesgue measure, and the Hausdorff dimension is a generalization of Euclidean (integer) dimension. Moreover, Hausdorff measure is a more refined notion of size than the Hausdorff dimension. Hausdorff's idea of measure is based on a Caratheodory approach of approximating X by countable covers. Let $f : \mathbb{R}^+ \to \mathbb{R}^+$ be an increasing continuous function such that $f(r) \to 0$ as $r \to 0$. The function f is usually referred to as a *dimension function*. Then the Hausdorff f-measure of a set $X \subset \mathbb{R}^n$ is defined as follows. For $\rho > 0$, a countable collection $\{B_i\}$ of Euclidean balls in $\mathbb{R}^n$

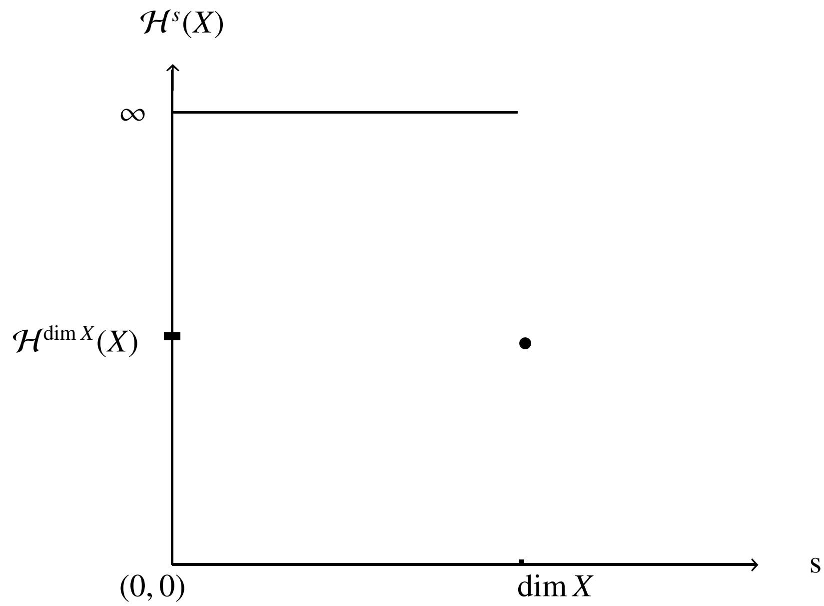

Figure 3.2: Graph of Hausdorff measure $\mathcal{H}^s(X)$ against the exponent s

with diameters $diam(B_i) \leq \rho$ such that $X \subset \bigcup_i B_i$ is called a ρ-cover for X. Define

$$\mathcal{H}^f_\rho(X) = \inf \left\{ \sum_i f(diam(B_i)) \ : \ \{B_i\} \text{ is a } \rho\text{-cover for } X \right\}$$

where the infimum is taken over all possible ρ-covers of X. The Hausdorff f-measure of X is defined as

$$\mathcal{H}^f(X) = \lim_{\rho \to 0^+} \mathcal{H}^f_\rho(X).$$

For the particular case $f(r) = r^s$, where $s > 0,$[1] $\mathcal{H}^f(X)$ is called the s-dimensional Hausdorff measure and is denoted by $\mathcal{H}^s(X)$.

Since the Hausdorff measure is defined in terms of the diameter of the covering sets, it is unchanged by restriction to closed, open, or convex sets. It is also unchanged by translations and rotations, but it is affected by scaling. Typically, the value $\mathcal{H}^s(X)$ jumps from infinity to zero as s increases. The value s at this discontinuity is called the Hausdorff dimension of the set X and is denoted by $\dim(X)$.

Formally

$$\dim(X) = \inf\{s \in \mathbb{R}^+ \ : \ \mathcal{H}^s(X) = 0\}$$

At the critical exponent $s = \dim X$, the quantity $\mathcal{H}^s(X)$ is zero, infinite, or strictly positive and finite. In the latter case, i.e.,

$$0 < \mathcal{H}^s(X) < \infty$$

the set X is said to be an s-set; see [49] for further details.

3.4 The Jarník–Besicovitch Theorem

Recall that if one considers the set $W(\psi)$ for an approximation function ψ of the form $\psi(q) = q^{-\tau}$, then $W(\psi)$ is the classical set of τ-approximable numbers $W(\tau)$. The convergence part of Khintchine's theorem implies that for any $\tau > 2$ the set $W(\tau)$ is of Lebesgue measure zero. Intuitively, one would expect the size of $W(\tau)$ to decrease as the rate of approximation increases, i.e., as τ increases. The following theorem, which is attributed to Jarník [56] and later independently to Besicovitch [57], allows us to distinguish between the size of the sets $W(\tau)$ expressed in terms of the Hausdorff dimension.

Theorem (Jarník–Besicovitch) 1 *For $\tau \geq 2$,*

$$\dim W(\tau) = \frac{2}{\tau}$$

The theorem confirms our intuition: the size of $W(\tau)$ expressed in terms of the Hausdorff dimension decreases as τ increases. In particular, $W(10)$ is indeed larger than $W(100)$.

It follows from the definition of the Hausdorff dimension and the Jarník–Besicovitch theorem that

$$\mathcal{H}^s(W(\tau)) = \begin{cases} 0 & \text{if} \quad s > 2/\tau \\ \\ \infty & \text{if} \quad s < 2/\tau \end{cases}$$

However, the Jarník-Besicovitch theorem gives no information regarding the s-dimensional Hausdorff measure of $W(\tau)$ at the critical value $s = \dim W(\tau)$. Thus, it is unable to distinguish between sets of the same dimension. Take, for example, the approximating functions

$$\psi_1(q) = \frac{1}{q^{10}} \qquad \text{and} \qquad \psi_2(q) = \frac{1}{q^{10}\log q} \tag{3.4}$$

It follows from the Jarník–Besicovitch theorem that

$$\dim W(\psi_1) = \dim W(\psi_2) = \frac{1}{5}$$

The final result, which distinguishes the exceptional sets of such kind as that presented in the above example, is called the Jarník theorem.

We have seen that the Jarník–Besicovitch theorem tells us the dimension of the sets $W(\tau)$. However, the following result reveals much more. It provides complete information concerning the Hausdorff f-measure of $W(\psi)$ and in particular allows us to determine the value of $\mathcal{H}^s(W(\psi))$ at the critical exponent $s = \dim(W(\psi))$.

Theorem (Jarník) 1 *Let ψ be an approximating function. Let f be a dimension function such that $q^{-1}f(q) \to \infty$ as $q \to \infty$ and $q^{-1}f(q)$ is decreasing. Then*

$$\mathcal{H}^f(W(\psi)) = \begin{cases} 0 & \text{if} \quad \sum_{q=1}^{\infty} q f(\psi(q)) < \infty \\ \infty & \text{if} \quad \sum_{q=1}^{\infty} q f(\psi(q)) = \infty \end{cases}$$

Clearly Jarník's theorem (see [58] for the original manuscript) can be regarded as the Hausdorff measure version of Khintchine's theorem. As with the latter, the divergence part constitutes the main substance. Notice that the case that $\mathcal{H}^f$ is comparable to the one-dimensional Lebesgue measure is excluded by the condition $q^{-1} f(q) \to \infty$ as $q \to 0$. Analogous to Khintchine's original statement, in Jarník's original statement the additional hypotheses that $q^2\psi(q)$ is decreasing, $q^2\psi(q) \to 0$ as $q \to \infty$, and that $q^2 f(\psi(q))$ is decreasing were assumed. The fact that these conditions are unnecessary is due to [55]. Thus, even in the simple case when $f(q) = q^s$ $(s \geq 0)$ and the approximating function is given by $\psi(q) = q^{-\tau}\log q$ $(\tau > 2)$, Jarník's original statement gives no information regarding the s-dimensional Hausdorff measure of $W(\psi)$ at the critical exponent $s = 2/\tau$. This is because $q^2 f(\psi(q))$ is not decreasing.

Recall that in the case that $\mathcal{H}^f$ is the standard s-dimensional Hausdorff measure $\mathcal{H}^s$ (i.e., $f(q) = q^s$), it follows from the definition of the Hausdorff dimension that

$$\dim W(\psi) = \inf\{s : \textstyle\sum_{q=1}^{\infty} q\,\psi(q)^s < \infty\}$$

In view of this, Jarník's zero–infinity law implies not only the Jarník–Besicovitch theorem, namely

$$\dim(W(\tau)) = 2/\tau \ \ (\tau \geq 2)$$

but also that

$$\mathcal{H}^{2/\tau}(W(\tau)) \ = \ \infty \ \ (\tau > 2)$$

Furthermore, the zero–infinity law allows us to discriminate between sets with the same dimension and even the same s-dimensional Hausdorff measure. For example, with $\tau \geq 2$ and $0 < \epsilon_1 < \epsilon_2$, consider the approximating functions

$$\psi_{\epsilon_i}(q) \ := \ q^{-\tau} \ (\log \ q)^{-\frac{\tau}{2}(1+\epsilon_i)} \qquad (i = 1, 2)$$

It is easily verified that for any $\epsilon_i > 0$

$$|(W(\psi_{\epsilon_i}))| = 0 \,, \quad \dim W(\psi_{\epsilon_i}) = 2/\tau \quad \text{and} \quad \mathcal{H}^{2/\tau}(W(\psi_{\epsilon_i})) = 0$$

However, consider the dimension function f given by

$$f(q) := q^{2/\tau}(\log q^{-1/\tau})^{\epsilon_1}$$

Then

$$\sum_{q=1}^{\infty} q\, f\,(\psi_{\epsilon_i}(q)) \ \asymp \ \sum_{q=1}^{\infty} (q\,(\log q)^{1+\epsilon_i-\epsilon_1})^{-1}$$

where, as usual, the symbol $\asymp$ denotes comparability (the quotient of the associated quantities is bounded from above and below by positive, finite constants). Hence, Jarník's zero–infinity law implies that

$$\mathcal{H}^f\,(W(\psi_{\epsilon_1})) \ = \ \infty \qquad \text{whilst} \qquad \mathcal{H}^f\,(W(\psi_{\epsilon_2})) \ = \ 0$$

Thus, the Hausdorff measure $\mathcal{H}^f$ does make a distinction between the sizes of the sets under consideration, unlike the s-dimensional Hausdorff measure.

3.4.1 A Concrete Example

Going back to the "concrete" approximating functions ψ_1 and ψ_2 given by (3.4), shows that both sets $W(\psi_1)$ and $W(\psi_2)$ have equal dimension, namely $1/5$. It is amazing that any difference between these sets can be found, considering that for large values of q, e.g., $q = 10^{100}$, $\psi_1(10^{100}) = \frac{1}{10^{1000}}$ whereas $\psi_2(10^{100}) = \frac{1}{10^{10000}\log(10)}$.
Let f be the dimension function given by

$$f(q) := q^{\frac{1}{5}}\left(\log q^{-1}\right)^{-1}$$

Therefore,

$$\begin{aligned}\sum_{q=1}^{\infty} q\, f(\psi_1(q)) &= \sum_{q=1}^{\infty}\left(q\ \log q^{10}\right)^{-1}\\ &\asymp \sum_{q=1}^{\infty}(q\ \log q)^{-1}\\ &= \infty\end{aligned}$$

whereas

$$\begin{aligned}\sum_{q=1}^{\infty} q\, f(\psi_2(q)) &= \sum_{q=1}^{\infty}\left(q\ (\log q)^{\frac{1}{5}}\ \log(q^{10}\ \log q\right)^{-1}\\ &\ll \sum_{q=1}^{\infty}\left(q\ (\log q)^{\frac{6}{5}}\right)^{-1}\\ &< \infty\end{aligned}$$

Hence, Jarník's zero–infinity law implies that

$$\mathcal{H}^f\left(W(\psi_1)\right) = \infty \qquad \text{whilst} \qquad \mathcal{H}^f\left(W(\psi_2)\right) = 0$$

3.5 The Linear Forms Theory

Results of the previous sections can be generalized to higher dimensions. These generalizations come in two distinct forms, namely, simultaneous and dual approximations. The two cases can be combined to give rise to linear form theory.

Let ψ be an approximating function. An $m \times n$ matrix $X = (x_{ij}) \in \mathbb{R}^{mn}$ is said to be ψ-*approximable* if the system of inequalities

$$\|q_1 x_{1i} + q_2 x_{2i} + \ldots + q_m x_{mi}\| < \psi(|\mathbf{q}|) \quad (1 \le i \le n) \tag{3.5}$$

is satisfied for infinitely many vectors $\mathbf{q} \in \mathbb{Z}^m \setminus \{\mathbf{0}\}$. Here $\|\cdot\|$ means distance to the nearest integer. For clarity, equation 4.16 may be expressed in the form

$$|q_1 x_{1i} + q_2 x_{2i} + \ldots + q_m x_{mi} - p_i| < \psi(|\mathbf{q}|) \quad (1 \le i \le n) \tag{3.6}$$

which is satisfied for infinitely many vectors $(\mathbf{p}, \mathbf{q}) = (p_1, \cdots, p_n, q_1, \cdots, q_m) \in \mathbb{Z}^n \times \mathbb{Z}^m \setminus \{\mathbf{0}\}$.

The system

$$q_1 x_{1i} + q_2 x_{2i} + \ldots + q_m x_{mi} \quad (1 \le i \le n)$$

of n real linear forms in m variables $q_1, \ldots, q_m$ will be written more concisely as $\mathbf{q}X$, where the matrix

$$X = (x_{ij}) := \begin{pmatrix} x_{11} & \ldots & x_{1n} \\ x_{21} & \ldots & x_{2n} \\ \vdots & & \vdots \\ x_{m1} & \ldots & x_{mn} \end{pmatrix}$$

is regarded as a point in $\mathbb{R}^{mn}$. It is easily seen that ψ-approximability is unaffected under translation by integer vectors, and we can therefore restrict attention to the unit cube $\mathbb{I}^{mn} := [-\frac{1}{2}, \frac{1}{2}]^{mn}$ as

$$\mathbb{R}^{mn} = \bigcup_{\mathbf{K} \in \mathbb{Z}^{mn}} (\mathbb{I}^{mn} + \mathbf{K})$$

The ψ-approximability in the linear forms setup is again rooted in the linear form version of the Dirichlet's theorem.

Theorem (Dirichlet for Vectors) 1 *Let N be a given natural number and let $X \in \mathbb{I}^{mn}$. Then there exists a non-zero integer $\mathbf{q} \in \mathbb{Z}^m$ with $1 \le |\mathbf{q}| \le N$ satisfying the system of inequalities*

$$\|q_1 x_{1i} + q_2 x_{2i} + \ldots + q_m x_{mi}\| < N^{-\frac{m}{n}} \quad (1 \le i \le n)$$

It can be easily deduced from the Dirichlet's theorem that

Corollary 2 *For any $X \in \mathbb{I}^{mn}$ there exist infinitely many integer vectors $\mathbf{q} \in \mathbb{Z}^m$ such that*

$$\|q_1 x_{1i} + q_2 x_{2i} + \ldots + q_m x_{mi}\| < |\mathbf{q}|^{-\frac{m}{n}} \quad (1 \le i \le n) \tag{3.7}$$

The right-hand side of (3.7) may be sharpened by a constant $c(m, n)$, but the best permissible values for $c(m, n)$ are unknown except for $m = n = 1$. It leads naturally to the notion of badly approximable points in $\mathbb{I}^{mn}$. A point $X \in \mathbb{I}^{mn}$ is said to be badly approximable if there exists a constant $C(X) > 0$ such that

$$\|\mathbf{q}X\| > C(X)|\mathbf{q}|^{-\frac{m}{n}}$$

for all $\mathbf{q} \in \mathbb{Z}^m \setminus \{\mathbf{0}\}$. Let $\mathbf{Bad}(m, n)$ denote the set of badly approximable points in $\mathbb{I}^{mn}$. A result of Schmidt [47] states that $\mathbf{Bad}(m, n)$ is a large set in the sense that it has maximal dimension, i.e.,

$$\dim \mathbf{Bad}(m, n) = mn$$

Let $W(m, n; \psi)$ denote the set of ψ-approximable points in $\mathbb{I}^{mn}$, i.e.,

$$W(m, n; \psi) := \{X \in \mathbb{I}^{mn} : \|\mathbf{q}X\| < \psi(|\mathbf{q}|) \text{ for i.m. } \mathbf{q} \in \mathbb{Z}^m \setminus \{\mathbf{0}\}\}$$

In this setup the points $X \in \mathbb{I}^{mn}$ are approximated by $(m-1)n$-dimensional hyperplanes. These play an analogous role to that of the rationals in the one-dimensional settings. The set $W(1, n; \psi)$ corresponds to simultaneous Diophantine approximation. In this setting the points $\mathbf{x} = (x_1, \ldots x_n) \in \mathbb{I}^n$ are approximated by the rational vectors $\mathbf{p}/q$, where $(\mathbf{p}, q) \in \mathbb{Z}^n \times \mathbb{N}$. More precisely, we deal with the inequalities

$$\begin{cases} |qx_1 - p_1| < \psi(q) \\ |qx_2 - p_2| < \psi(q) \\ \qquad \cdots\cdots \\ |qx_n - p_n| < \psi(q) \end{cases}$$

The set $W(m, 1; \psi)$ corresponds to the dual Diophantine approximation. In the dual approximation, instead of approximated by rational points, one considers the closeness of points

$\mathbf{x} := (x_1, \ldots, x_m) \in \mathbb{I}^m$ to rational hyperplanes given by the equation

$$\mathbf{q} \cdot \mathbf{x} = p \ \text{ with } \ (p, \mathbf{q}) \in \mathbb{Z} \times \mathbb{Z}^m$$

i.e., we work with the inequality

$$|q_1 x_1 + q_2 x_2 + \ldots + q_m x_m - p| \ < \ \psi(|\mathbf{q}|)$$

3.5.1 Khintchine–Groshev Theorem

The main Lebesgue result in the linear form settings is the Khintchine–Groshev theorem, which gives an elegant answer to the question of the size of the set $W(m, n; \psi)$. The result links the measure of the set to the convergence or otherwise of a series that depends only on the approximating function and is the template for many results in the field of metric number theory. It provides a complete answer to the question of Lebesgue measure of the ψ-approximable points. The following statement is due to Groshev [46] and extends Khintchine's simultaneous result [59] to the dual form case.

Theorem (Khintchine-Groshev) 1 *Let ψ be an approximating function. Then*

$$|W(m, n; \psi)|_{mn} = \begin{cases} 0 & \text{if} \quad \sum_{r=1}^{\infty} r^{m-1} \psi(r)^n < \infty \\ 1 & \text{if} \quad \sum_{r=1}^{\infty} r^{m-1} \psi(r)^n = \infty \end{cases}$$

where $|W(m, n; \psi)|_{mn}$ denotes the mn-dimensional Lebesgue measure of the set $W(m, n; \psi)$. The proof of the convergence case of the Khintchine–Groshev theorem is easily established by a straightforward application of the Borel–Cantelli lemma and is free from any assumption on ψ. The divergence part constitutes the main substance of the Khintchine–Groshev theorem and requires the monotonicity assumption on the function ψ. It is worth mentioning that in the original statement of the Khintchine–Groshev theorem [46, 51, 59] the stronger hypothesis that $r^{\max(1, m-1)} \psi(r)^n$ is monotonic was assumed. The fact that this assumption is unnecessary is due to [52]. The case when the sum diverges was proved by Groshev using Fourier analysis. It can also be proved using probabilistic ideas that rely on the pairwise statistical independence of sets associated with pairs of integer vectors that are linearly independent (see [60] for details).

Let us return briefly to the set $\mathbf{Bad}(m,n)$ of badly approximable points. It follows from the definition of $\mathbf{Bad}(m,n)$ and $W(m,n;\psi)$ that

$$\mathbf{Bad}(m,n) \subset \mathbb{I}^{mn} \setminus W(m,n;\psi)$$

In view of the Khintchine–Groshev theorem, the set on the right-hand side is of Lebesgue measure zero, and thus

$$|\mathbf{Bad}(m,n)|_{mn} \;=\; 0$$

In the one-dimensional case ($m = n = 1$), it is well known that the monotonicity hypothesis in the Khintchine–Groshev theorem is absolutely crucial (see §3.2). In other words, the Khintchine–Groshev theorem is false without the monotonicity hypothesis, and the Duffin–Schaeffer [53] conjecture provides an appropriate alternative statement. Beyond the one-dimensional setting, the problem of removing monotonicity on the approximating function of the Khitchine–Groshev theorem is fully settled.

Theorem 3 *Let* $\psi : \mathbb{N} \to \mathbb{R}^+$ *and* $mn > 1$. *Then*

$$|W\,(m,n;\psi)\,|_{mn} \;=\begin{cases} 0 & \text{if} \quad \displaystyle\sum_{r=1}^{\infty} r^{m-1}\psi(r)^n < \infty \\[2ex] 1 & \text{if} \quad \displaystyle\sum_{r=1}^{\infty} r^{m-1}\psi(r)^n = \infty \end{cases}$$

The proof of Theorem 3 is attributed to various authors for different values of m. For $m = 1$, Theorem 3 was proved by Gallagher [61]. For $m = 2$, Theorem 3 was recently proved by Beresenevich and Velani [62]. For $m \geq 3$, it can be derived from Schmidt [63, Theorem 2] or Sprindzŭk's [50, §1.5, Theorem 15].

Note that the Lebesgue theory so far discussed is rigid in the sense that it only tells us measure as zero or one. As a result the sets that obey zero–one laws always involve exceptional sets of measure zero. The Lebesgue theory does not tell us anything more about the size of these exceptional sets, although it is clear that it should depend on the choice of the approximating function.

With reference to the system of linear form, the Khintchine–Groshev theorem shows that if the approximating function ψ decreases sufficiently quickly so that the sum converges, the corresponding set of ψ -approximable points is of zero mn-dimensional Lebesgue measure. In this case we cannot obtain Any further information regarding the size of $W(m,n;\psi)$ in terms of Lebesgue measure. Intuitively, the size of $W(m,n;\psi)$ should decrease as the speed of approximation increases. The following theorem allows us to distinguish between sizes of the sets $W(m,n;\psi)$ expressed in terms of the Hausdorff dimension.

3.5.2 Multidimensional Jarník–Besicovitch Theorem

Dodson et al. [64] showed that the Hausdorff dimension of ψ-approximable sets is related to the lower order at infinity of the function $\frac{1}{\psi}$. The lower order (at infinity) $\lambda(g)$ of a function $g : \mathbb{R}^+ \to \mathbb{R}^+$ is defined by

$$\lambda(g) := \liminf_{r\to\infty} \frac{\log g(r)}{\log r}$$

and indicates the growth of g "near" infinity; it is non-negative for an increasing function.

Theorem (Dodson) 1 *Let ψ be an approximating function and let λ be the lower order at infinity of $\frac{1}{\psi}$. Then*

$$\dim W(m,n;\psi) = \begin{cases} (m-1)n + \frac{m+n}{\lambda+1} & \text{if} \quad \lambda > \dfrac{m}{n} \\ \\ mn & \text{if} \quad \lambda \leq \dfrac{m}{n} \end{cases}$$

This theorem can be regarded as the Hausdorff dimension version of the Khintchine–Groshev theorem, in which the Lebesgue measure is replaced by the Hausdorff dimension and the volume sum is replaced by the lower order at infinity. Intuitively, the more rapid the approximation, the larger the lower order λ and so smaller the dimension of the set $W(m,n;\psi)$. It follows from the definition of the Hausdorff dimension and Dodson's theorem that

$$\mathcal{H}^s(W(m,n;\psi)) = \begin{cases} 0 & \text{if} \quad s > \dim W(m,n;\psi) \\ \\ \infty & \text{if} \quad s < \dim W(m,n;\psi) \end{cases}$$

However, Dodson's theorem gives no information regarding the s-dimensional Hausdorff measure of $W(m,n;\psi)$ at the critical exponent $s = \dim W(m,n;\psi)$. Thus, it is unable to distinguish between the sets of the same dimension. In the simultaneous approximation case Dodson's theorem was essentially proved by Jarník [58]. In fact, Jarník obtained the Hausdorff dimension result as a corollary of a Hausdorff measure version of Khintchine's theorem. Dickinson and Velani [65] obtained the Hausdorff measure analogue of the Khintchine–Groshev theorem.

Theorem (Dickinson and Velani) 1 *Let ψ be an approximating function and let f be a dimension function such that $r^{-mn}f(r) \to \infty$ as $r \to 0$ and $r^{-mn}f(r)$ is decreasing. Furthermore, suppose that $r^{-(m-1)n}f(r)$ is increasing. Then,*

$$\mathcal{H}^f(W(m,n;\psi)) = \begin{cases} 0 & \text{if} \quad \sum_{r=1}^{\infty} f(\psi(r))\psi(r)^{-(m-1)n} r^{m-1} < \infty \\ \\ \infty & \text{if} \quad \sum_{r=1}^{\infty} f(\psi(r))\psi(r)^{-(m-1)n} r^{m-1} = \infty \end{cases}$$

As in all the measure theoretic statements, the divergence part constitutes the main substance of the theorem. The case when $\mathcal{H}^f$ is comparable with the mn-dimensional Lebesgue measure is excluded by the condition $r^{-mn}f(r) \to \infty$ as $r \to 0$. As with the Jarník theorem in one-dimensional settings, the zero–infinity law of Theorem of Dickinson and Velani allows us to discriminate between sets with the same dimension and even the same s-dimensional Hausdorff measure.

3.5.3 A Unified Statement

The Khintchine–Groshev theorem and Theorem of Dickinson and Velani can be combined to obtain a single unifying statement.

Theorem 4 *Let ψ be an approximating function. Let f be a dimension function such that $r^{-mn}f(r)$ is monotonic and $r^{-(m-1)n}f(r)$ is increasing. Then,*

$$\mathcal{H}^f\,(\,W(m,n;\psi)\,) = \begin{cases} 0 & \text{if} \quad \displaystyle\sum_{r=1}^{\infty} f(\psi(r))\psi(r)^{-(m-1)n} r^{m-1} < \infty \\ \\ \mathcal{H}^f(\mathbb{I}^{mn}) & \text{if} \quad \displaystyle\sum_{r=1}^{\infty} f(\psi(r))\psi(r)^{-(m-1)n} r^{m-1} = \infty \end{cases}$$

For monotonic approximating functions Theorem 4 provides a complete measure theoretic description of $W(m,n;\psi)$. In the case when $f(r) := r^{mn}$ the Hausdorff measure $\mathcal{H}^f$ is simply the standard Lebesgue measure. The condition that $r^{-mn}f(r)$ is monotonic is not restrictive and gives an infinite measure statement (a multidimensional analogue of Jarník's theorem) when $r^{-mn}f(r) \to \infty$ as $r \to 0$. It is notable that the underlying proof of both the statements is different. However, in view of the Mass Transference Principle for linear forms, recently established in [66], one obtains the Hausdorff measure results from the Lebesgue measure statements. In other words

$$\text{Khintchine–Groshev Theorem} \quad \Longrightarrow \quad \text{Jarník's Theorem}$$

Chapter 4

DoF of MIMO X Channel with Constant Channel Gains

In this chapter the interference alignment technique is deployed in detail to characterize the total DoF of MIMO $(K \times 2, M)$ and $(2 \times K, M)$ X channels both with real and complex channel realization. Techniques from Chapter 3 are used to provide a new set of Diophantine approximation theorems to enable the possibility of simultaneous joint decoding at receivers.

4.1 System Model

4.1.1 Notation

Throughout this chapter, boldface upper-case letters, e.g., $\mathbf{H}$, are used to represent matrices. Matrix elements will be shown in brackets, e.g., $\mathbf{H} = [h_{i,j}]$ for a set of values i, j. Vectors are shown using boldface italic lower-case letters, e.g., $\boldsymbol{v}$. Vector elements are shown inside parentheses, e.g., $\boldsymbol{v} = (v_1, v_2, ..., v_i)$ for a set of values i. The transpose and conjugate transpose of a matrix $\mathbf{A}$ will be represented as $\mathbf{A}^t$ and $\mathbf{A}^\dagger$ respectively. The general transmitted signal from the kth antenna of transmitter i desired to be decoded at receiver j is represented by $x_k^{i,j}$. At each antenna of transmitters in the X channel, a linear combination of all desired messages for different receivers will be transmitted. To simplify notation it is assumed that $x_k^i = \sum_j \beta_j x_k^{i,j}$, where β_j is the weight of message $x_k^{i,j}$ in the linear combination. The transmitted vector signal at transmitter i will be represented as $\boldsymbol{x}^i = (x_1^i, x_2^i, ..., x_k^i)^t$ for a set of values k. We use single superscript labelling for the indices of transmitters and receivers, for example, $\boldsymbol{z}^i$ represents the noise vector at the receiver i. Single subscripts are used for the antenna labelling unless otherwise stated; for example, y_j^i represents the received signal at the jth antenna of receiver i. The superscript pair i, j assigns the variable from transmitter i to receiver j, and similarly the subscript pair l, n represents the variable from antenna l to antenna n. For example, $h_{l,n}^{i,j}$ represents the channel between the lth antenna of transmitter i and the nth antenna of the receiver j.

4.1.2 K-Transmitter, 2-Receiver M Antenna X Channel

A constant fully connected K-transmitter, 2-receiver MIMO Gaussian X network is considered. This channel is used to model a communication network with K transmitters and two receivers. Each transmitter is equipped with M antennas and wishes to communicate with both receivers, transmitting a dedicated message to each of them. Each of the receivers is also equipped with M antennas. All transmitters share a common bandwidth. The channel outputs at the receivers are characterized by the following input–output relationship:

$$\boldsymbol{y}^i = \mathbf{H}^{1,i}\boldsymbol{x}^1 + \mathbf{H}^{2,i}\boldsymbol{x}^2 + ... + \mathbf{H}^{K,i}\boldsymbol{x}^K + \boldsymbol{z}^i$$

where $i \in \{1, 2\}$ is the receiver index, $k \in \{1, 2, ..., K\}$ is the user index, $\boldsymbol{y}^i = (y_1^i, y_2^i, ..., y_M^i)^t$ is the $M \times 1$ output signal vector of ith receiver, $\boldsymbol{x}^j = (x_1^j, x_2^j, ..., x_M^j)^t$ is the $M \times 1$ input vector signal of the jth transmitter, $\mathbf{H}^{j,i} = [h_{l,n}^{j,i}]$ is the $M \times M$ channel matrix between transmitter j and receiver i, where $h_{l,n}^{j,i}$ specifies the channel gain from the lth antenna of jth transmitter to the nth antenna of the receiver i, and $\boldsymbol{z}^i = (z_1^i, z_2^i, ..., z_M^i)^t$ is $M \times 1$ Additive White Gaussian Noise (AWGN) vector at the receiver i. All noise terms are assumed to be independent and identically distributed (i.i.d.) zero mean, unit variance Gaussian random variables. It is assumed that each transmitter is subject to an average power constraint P:

$$\mathbb{E}[(\boldsymbol{x}^j)^\dagger(\boldsymbol{x}^j)] \leq P$$

where $\mathbb{E}[x]$ represents the expectation of the random variable x. As mentioned earlier in the notation section of this chapter, the transmitted signal from the kth antenna of transmitter i desired to be decoded at receiver j is represented by $x_k^{i,j}$. At each antenna of each transmitters in the X channel, a linear combination of all desired messages for different receivers will be transmitted. It is assumed that $x_k^i = \sum_j \beta_j x_k^{i,j}$, where β_j is the weight of message $x_k^{i,j}$ in the linear combination.

Let $P_e^{j,i}$ denote the probability of error for a message sent by transmitter j and received at receiver i, i.e.,

$$P_e^{j,i} = Pr\{W^{j,i} \neq \hat{W}^{j,i}\}$$

where $W^{j,i}$ is the message sent by transmitter j to receiver i with the rate of $R^{j,i}$ and $\hat{W}^{j,i}$ is decoded message from this transmission. For a given power constraint P, a rate region $R(P)$ is determined by rates $R^{j,i}$. The closure of the set of all achievable rate tuples is called the capacity region of the channel with power constraint P and is denoted by $C(P)$. The notion of DoF is defined next.

Definition 2 *To an achievable rate tuple $R(P) \in C(P)$ one can correspond an achievable DoF of $d^{j,i}$ provided that*

$$R^{j,i} = \frac{1}{2} d^{j,i} \log_2(P) + o(\log_2(P))$$

The set of all achievable DoF tuples is called the DoF region and is denoted by $\mathscr{D}$.

Definition 3 *The maximum sum rate or sum capacity of the K-transmitter, 2-receiver MIMO Gaussian X channel is defined as*

$$\boldsymbol{C}_{\Sigma}(P) = \max_{R^{j,i} \in C(P)} \sum_{i=1}^{2} \sum_{j=1}^{K} R^{j,i}$$

The maximum achievable sum DoF (or simply channel total DoF) is defined as

$$D = \max_{d^{j,i} \in \mathcal{D}} \sum_{i=1}^{2} \sum_{j=1}^{K} d^{j,i}$$

In sequel, the $(K \times 2, M)$ X channel refers to constant channel gain, K-transmitter, 2-receiver MIMO X channel with M antennas at each transmitter/receiver.

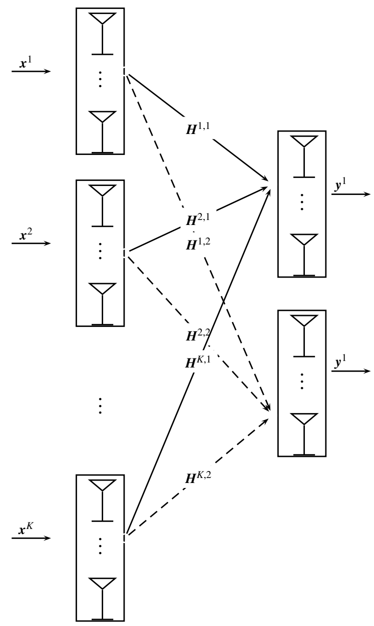

Figure 4.1: $K \times 2$, M antenna X channel

4.1.3 2-Transmitter, K-Receiver M Antenna X Channel

A constant fully connected 2-transmitter, K-receiver MIMO Gaussian X network is considered. Each of the transmitters and receivers are equipped with M antennas. The schematic of this channel can be found in Figure 4.2. The channel outputs at the receivers are characterized by the following input–output relationship:

$$\boldsymbol{y}^i = \mathbf{H}^{1,i}\,\boldsymbol{x}^1 + \mathbf{H}^{2,i}\boldsymbol{x}^2 + \boldsymbol{z}^i$$

where $i \in \{1, 2, ..., K\}$ is the receiver index and $\boldsymbol{z}^i = (z^i_1, z^i_2, ..., z^i_M)^t$ is the $M \times 1$ AWGN vector at the receiver i. It is assumed that all noise terms are i.i.d. zero mean unit variance Gaussian random variables and that each transmitter is subject to an average power constraint P. The $\boldsymbol{x}^j = (x^j_1, x^j_2, ..., x^j_K)^t$ is the transmitted message from transmitter j to all receivers.

Similar to the $K \times 2$, MIMO X channel, sum capacity and DoF region for $2 \times K$, MIMO X channels can be defined. In sequel, the $(2 \times K, M)$ X channel refers to constant channel gain, 2-transmitter, K-receiver MIMO X channel with M antennas at each transmitter/receiver.

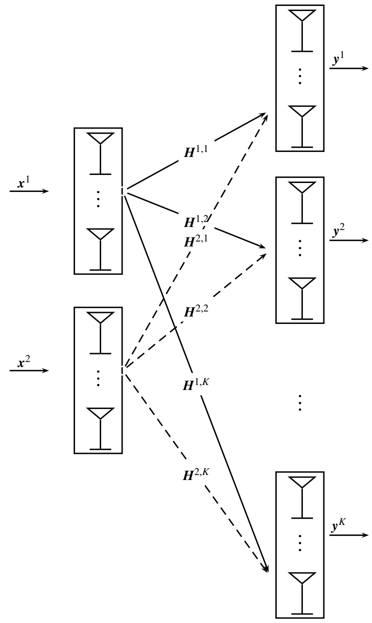

Figure 4.2: $2 \times K$, M antenna X channel

4.2 Main Contribution and Discussion

4.2.1 Main Result

In this thesis, the total DoF of the following channels are characterized:

1. $(2 \times K, M)$ X channel with constant real/complex channel realizations
2. $(K \times 2, M)$ X channel with constant real/complex channel realizations

It is observed that the duality/reciprocity holds for the DoF of this class of X channels, i.e., if the role of transmitters is interchanged with that of receivers the total DoF of the network will be conserved. The technique that is used in this thesis, named layered interference alignment, benefits from a linear pre-coding similar to vector interference alignment in conjunction with a number theoretic technique similar to the real alignment using rational dimensions at the transmitter. At the receiver a new mathematical tool was first introduced to empower the ability to use joint processing and mutual decoding among the receiver antennas to achieve the total fractional DoF of each desired message. The main results can be stated as follows:

Theorem 5 *The total DoF of $(K \times 2, M)$ X channel with real and time invariant channel coefficients is $\frac{2KM}{K+1}$ for almost all channel realizations.*

Theorem 6 *The total DoF of the $(2 \times K, M)$ X channel with real and time invariant channel coefficients is $\frac{2KM}{K+1}$ for almost all channel realizations. This implies that when the base for comparison is the DoF, $(2 \times K, M)$ and $(K \times 2, M)$ X channels are dual/reciprocal.*

Theorem 7 *The total DoF of the $(2 \times K, M)$ X channel and its dual, which is the $(K \times 2, M)$ X channel with complex and time invariant channel coefficients, is $\frac{4KM}{K+1}$ for almost all channel realizations. This is twice that of the same channel with real channel gains. (The DoF for complex channel realizations should be defined as half of its definition for real cases, since complex transmission uses two dimensions for each transmission. This implies that the total DoF per transmit dimension is the same as real channel realization and is $\frac{2KM}{K+1}$.)*

It is noteworthy that two novel number theoretical theorems are introduced to prove the above results; these will be discussed in Sections 4.4 and 4.7.

4.2.2 Layered Interference Alignment

As described earlier, both previously known interference alignment methods, real and vector interference alignment, can be used to reduce the effect of interference at the receiver side, but there are still several cases where neither of them can achieve the total DoF of the system. In this dissertation, a new powerful tool, layered interference alignment, is introduced. It contains ingredients from both real and vector interference alignment techniques together with a novel number theoretic *Joint Processing/Simultaneous Decoding* scheme at the receiver side. Alignment of interfering signals is performed in two levels. First, using a vector alignment type of constellation design, transmit directions are chosen in such a way that the dimension occupied by the interference at all receivers is minimized. Second, similar to the real interference alignment scheme, data streams are modulated at integer values and multiplied by real numbers derived from channel coefficients. This allows further reduction in the interference subspace.

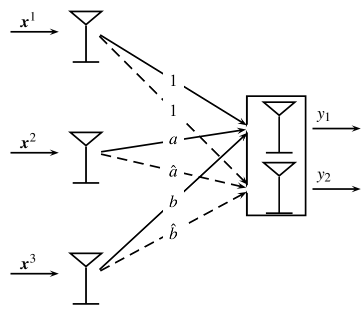

Figure 4.3: SIMO multiple access channel

Each received antenna observes a linear combination of the intended signals and aligned interfering ones in a single stream. In order to decode the desired message, all received antennas will participate in a joint processing scheme, based on a new Khintchine–Groshev type of theorem for badly approximable numbers, which will be proved in Section 4.4.1.

4.3 Preliminaries

In this section, first the signal design for encoding and decoding using interference alignment is introduced, and the required definitions that are used to prove the main results are gone through. Next, the measure theoretic results of Khintchine–Groshev type theorems are used to analyze the performance signal interference alignment in rational dimensions. Then, Diophantine approximation similar to real alignment is used to characterize the total DoF of $K \times M$ SISO X channels. Finally, at the end of this section the notation of *All most cases*, which is a part of all of these results, is discussed in detail.

4.3.1 Main Ideas and Basic Examples

To clarify basic ideas, we rely on a simple example and provide only rough reasoning for rationality of the schemes. Unless otherwise stated, the following assumptions are in place throughout this chapter.

Example 1 (Single-Input Multiple-Output (SIMO) Multiple Access Channel) *A multiple access channel with three single antenna users and a 2-antenna receiver is shown in Figure 4.3. The channel can be modelled as*

$$\begin{cases} y_1 = x^1 + ax^2 + bx^3 + z_1 \\ y_2 = x^1 + \hat{a}x^2 + \hat{b}x^3 + z_2 \end{cases} \tag{4.1}$$

where all channel parameters are constant real numbers.

Since the capacity region of this channel is fully characterized, it can be easily shown that the total DoF of 2 is achievable. Vector interference alignment falls short of achieving this DoF, as transmitters are equipped with a single antenna. The naive application of real interference alignment results in the same conclusion. To see this, let us assume that all three users communicate with the receiver using a single data stream. The data streams are modulated by the constellation $\mathcal{U} = A(-Q, Q)_{\mathbb{Z}} = \{$all integers between $-Q$ and $Q\}$, where A is a factor controlling the minimum distance of the received constellation.

The received constellation, which is a set of points in a two-dimensional space, consists of points $(v, \hat{v})$ such that $v = A(u^1 + au^2 + bu^3)$ and $\hat{v} = A(u^1 + \hat{a}u^2 + \hat{b}u^3)$, where u^i 's are members of $\mathcal{U}$. Let us choose two sets of distinct points $(v_1, \hat{v}_1)$ and $(v_2, \hat{v}_2)$ in the received constellation. The Khintchine–Groshev theorem provides us a lower bound on any linear combination of integers. It also provides some bound on the distance between any integer vector and the linear combination of rationally independent vectors. Using the theorem, one can obtain $d_{\min} \approx \frac{A}{Q^2}$, where $d_{\min}$ is the minimum distance in the received constellation.

By using the noise removal definition (Def. 1) and choosing the unity variance for the Gaussian noise, the noise can be removed if $d_{\min} = 1$. Hence, it is sufficient to have $A \approx Q^2$. In a noise-free environment, each received antenna can decode the three messages if there is a one-to-one map from the received constellation to the transmit constellations. Mathematically, one can satisfy this condition by enforcing the following:

Remark 4 (Separability Condition) *Each received antenna is able to decode all three messages if the channel coefficients associated with that antenna are rationally independent. In the above multiple access channel, for instance, the receiver can decode all messages by using the* *signal from the first antenna if* $u^1 + au^2 + bu^3 = 0$ *has no non-trivial solution in integers for* u^1, u^2 *and* u^3.

If the above condition is assumed and individual signals from each antenna used, the receiver is able to decode all messages. This is in fact the main drawback of real interference alignment, as it cannot exploit the availability of multiple antennas.

User i's rate is $R^i = \log(2Q - 1)$. Because of the power constraints, $P = A^2Q^2$. It was shown earlier in this chapter that $A \approx Q^2$. Therefore, $P \approx Q^6$. Hence,

$$d^i = \lim_{P\to\infty} \frac{R^i}{0.5 \log P} = \frac{1}{3} \tag{4.2}$$

If all three messages are decoded, the achievable DoF for this channel would be 1, which is not desired, as the total DoF is 2. In [24] and as discussed earlier, the real interference alignment can achieve the total DoF for multiple access channel and also two-user X channels both in the single antenna setup, but this scheme cannot solve the general MIMO setup.

A new alignment scheme called layered interference alignment is proposed to achieve the total DoF of this channel and a class of MIMO channels. This technique, in general, combines vector and real interference alignment techniques in a subtle way to enjoy the benefit of multiple transmit antennas at both transmitter and receiver sides. The multiple access channel considered in this section has no room for vector alignment. This helps us to understand the difference between the real and the layered interference alignment. Concretely, the layered interference alignment adds the joint processing of the received signals at the receiver side to the original real alignment by incorporating a new Khintchine–Groshev type theorem for approximating some badly approximable number sets. This theorem bounds the $d_{\min}$ based on the size of the input constellation and the number of antennas, in the same fashion that was discussed in [67]. These results are backed up by Theorem 8, which will be discussed in detail in section 4.4. To use these mathematical results one must provide an algorithm at receivers for simultaneous decoding.

Definition 4 (Joint processing of received data streams) *Each receiver first normalizes the received data streams in order to have the unity coefficient for a specified favourite message* *at all received antennas. Then it uses the results of Theorem 8 to simultaneously decode each message from all received streams at each of all the M antennas. Then it applies the same procedure for all other desired messages.*

In the multiple access channel example, joint decoding is employed at both received antennas. User i's rate is $R^i = \log(2Q-1)$; in the presence of the power constraint, we have $P = A^2Q^2$. Applying Theorem 8 and satisfying the noise removal assumption, we will have $A \approx Q^{0.5}$. User i's rate is $R^i = \log(2Q-1)$ because of the power constraint, and we have $P = A^2Q^2$. Applying Theorem 8 and satisfying the noise removal assumption will give $A \approx Q^{0.5}$. User i's rate is $R^i = \log(2Q-1)$. Because of the power constraint we have $P = A^2Q^2$. Applying Theorem 8 and satisfying the noise removal assumption will give $A \approx Q^{0.5}$. Therefore, $P \approx Q^3$. So

$$d^i = \lim_{P\to\infty} \frac{R^i}{0.5\log P} = \frac{2}{3} \tag{4.3}$$

Using the above method to decode each of the three messages, each of which has $\frac{2}{3}$, gives the total DoF of 2, which is the desired result. In the rest of this chapter, we incorporate layered interference alignment in its full potential, i.e., having the vector and the real interference alignment together with joint processing, to achieve the total DoF of $(K \times 2, M)$ and $(2 \times K, M)$ X channels.

Remark 5 (Complex Coefficients) *Despite the multiple access channel, it can be easily seen that the total DoF of the X channel with complex coefficients cannot be derived by pairing. In fact, a simple extension of the coding proposed in this thesis results in the total DoF of this channel [68]. In this case, using layered interference alignment requires a new joint processing bound, which was discussed separately in Section 4.7. This new theorem leaves the encoding and decoding methods intact and provides the required analytic to analyze the performance of the layered interference alignment for the constant complex number channel gain realizations. It will be observed that this powerful technique will achieve the total DoF of $\frac{4KM}{K+1}$ for both $(K \times 2, M)$ and $(2 \times K, M)$ X channels with constant complex channel gains. (This is twice the DoF of the same channels with real channel coefficients, due to use of two dimensions for complex number transmission.)*

It must be emphasized that to benefit from the properties of Khintchine–Groshev type theorems, similar to the real alignment [24], all the results and proofs presented in this thesis are based on the separability condition and Khintchine–Groshev type theorems. Therefore, all results are valid for almost all channel realizations.

4.4 A new Simultaneous Diophantine Approximation

Let $\psi : \mathbb{R}^+ \to \mathbb{R}^+$ be a real positive decreasing function with $\psi(r) \to 0$ as $r \to \infty$. Such a function will be referred to as an *approximating* function. Let $W(m, n; \psi)$ be the set of $X \in \mathbb{I}^{mn} := [-\frac{1}{2}, \frac{1}{2}]^{mn}$ such that the system of inequalities

$$|q_1 x_{1i} + \cdots + q_m x_{mi} - p| < \psi(|\mathbf{q}|) \quad 1 \leq i \leq n \tag{4.4}$$

is satisfied for infinitely many $(p, \mathbf{q}) \in \mathbb{Z} \times \mathbb{Z}^m \setminus \{\mathbf{0}\}$.

Here and throughout, the system $q_1 x_{1i}+, \ldots, +q_{mi}$ of n linear forms in m variables will be written more concisely as $\mathbf{q}X$, and $|\mathbf{q}|$ denotes the supremum norm of the integer vector $\mathbf{q}$.

The set $W(m, n; \psi)$ is a hybrid of the classical set in which the distance to the nearest integer is allowed to vary from one linear form to the other. In this situation it is the same for all the linear forms. Sets of similar nature have been studied by Hussain and his collaborators in [69], [70] and [71]. The Khintchine–Groshev type result is proved for $W(m, n; \psi)$. The results throughout this section crucially depend upon the choices of m and n, similar to the above-mentioned thesis and unlike the classical sets.

Theorem 8 *Let $m + 1 > n$ and ψ be an approximating function; then*

$$|W(m, n; \psi)|_{mn} = \begin{cases} 0 & \text{if } \sum_{r=1}^{\infty} \psi(r)^n r^{m-n} < \infty \\ 1 & \text{if } \sum_{r=1}^{\infty} \psi(r)^n r^{m-n} = \infty \end{cases}$$

The convergence half follows from the Borel–Cantelli lemma by construction of a suitable cover for the set $W(m, n; \psi)$. It does not rely on the choices of m and n, and it is free from monotonic assumption on the approximating function. The divergence case can be proved by using the similar arguments as in [69].

4.4.1 Proof of The Convergence Case of Theorem 8

The convergence case follows from applying the Borel–Cantelli lemma after making use of the natural cover of the set $W(m,n;\psi)$. The resonant sets are defined as

$$R_q = \{X \in \mathbb{I}^{mn} : \mathbf{q}X - p = 0\} \tag{4.5}$$

Thus, the resonant sets are $(m-1)n$-dimensional hyperplanes passing through the point p. The set $W(m,n;\psi)$ can be written as a lim sup set using the resonant sets in the following way.

$$W(m,n;\psi) = \bigcap_{N=1}^{\infty} \bigcup_{r>N} \bigcup_{R_q : |\mathbf{q}|=r} B\left(R_q, \psi(|\mathbf{q}|)\right)$$

where

$$B\left(R_q, \psi(|\mathbf{q}|)\right) = \left\{X \in I^{mn} : \operatorname{dist}\left(X, R_q\right) \leq \frac{\psi(|\mathbf{q}|)}{|\mathbf{q}|}\right\}$$

Figure 4.4 corresponds to the values $m=2, n=1$. The resonant set $R_{\mathbf{q}}$ is a line $q_1x + q_2y - p = 0$, intercepting the x and y axes at $\frac{p}{q_1}$ and $\frac{p}{q_2}$, respectively. The set $B\left(R_{\mathbf{q}}, \psi(|\mathbf{q}|)\right)$ is the $\frac{\psi(|\mathbf{q}|)}{|\mathbf{q}|}$ neighbourhood of $R_{\mathbf{q}}$.

Thus, for each $N \in \mathbb{N}$ the family

$$\left\{ \bigcup_{R_q : |\mathbf{q}|=r} B\left(R_q, \psi(|\mathbf{q}|)\right) : r = N, N+1, \ldots \right\}$$

is a cover for the set $W(m,n;\psi)$. Now, for each resonant set $R_{\mathbf{q}}$, let $\Delta(\mathbf{q})$ be a collection of mn-dimensional closed hypercubes C with disjoint interiors and side length comparable with $\psi(|\mathbf{q}|)/|\mathbf{q}|$ and diameter at most $\psi(|\mathbf{q}|)/|\mathbf{q}|$ such that

$$C \cap \bigcup_{R_{\mathbf{q}} : |\mathbf{q}|=r} B\left(R_{\mathbf{q}}, \psi(|\mathbf{q}|)\right) \neq \emptyset$$

and

$$B\left(R_{\mathbf{q}}, \psi(|\mathbf{q}|)\right) \subset \bigcup_{C \in \Delta(\mathbf{q})} C$$

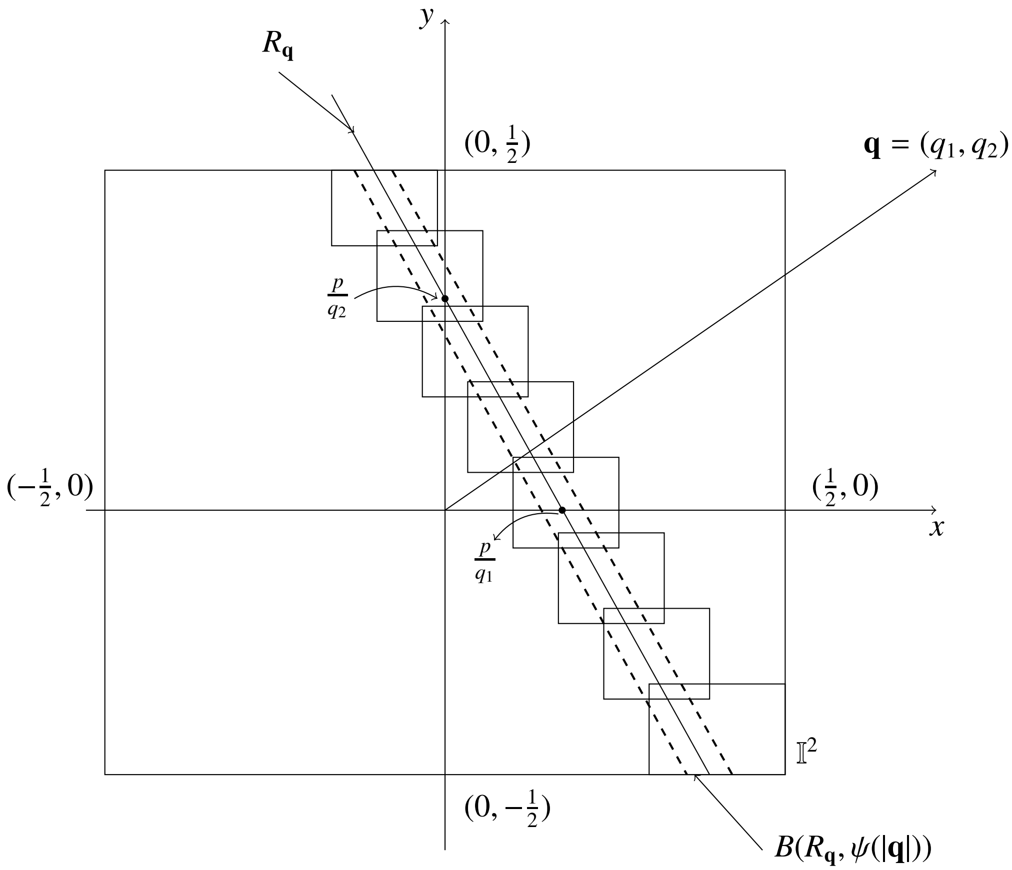

Figure 4.4: The resonant set $R_{\mathbf{q}}$ is a line for $m = 2$ and $n = 1$

Then

$$\#\Delta(\mathbf{q}) \ll (\psi(|\mathbf{q}|)/|\mathbf{q}|)^{-(m-1)n}$$

where # denotes cardinality. Note that

$$\begin{aligned} W(m,n;\psi) &\subset \bigcup_{r>N} \bigcup_{R_q:|\mathbf{q}|=r} \Delta\left(R_{\mathbf{q}}, \Psi(|\mathbf{q}|)\right) \\ &\subset \bigcup_{r>N} \bigcup_{\Delta(\mathbf{q}):|\mathbf{q}|=r} \bigcup_{C\in\Delta(\mathbf{q})} C \end{aligned}$$

Hence,

$$\begin{aligned} |W(m,n;\psi)| &\leq \sum_{r>N} \sum_{\Delta(\mathbf{q}):|\mathbf{q}|=r} \sum_{C\in\Delta(\mathbf{q})} |C| \\ &\ll \sum_{r>N} r^m \left(\frac{\psi(r)}{r}\right)^{mn} \left(\frac{\psi(r)}{r}\right)^{-(m-1)n} \\ &= \sum_{r>N} r^{m-n}\psi(r)^n \end{aligned}$$

since the sum $\sum_{r=1}^{\infty} \psi(r)^n r^{m-n}$ is convergent, which gives zero Lebesgue measure by Borel–Cantelli lemma.

4.4.2 Proof of The Divergence Case of Theorem 8

For the divergence case the ubiquity theorem Theorem 1 in [72, Theorem 1] is used, and to establish ubiquity two technical lemmas (Lemma 1 and Lemma 2) are needed. The work is similar to [69]; therefore, I only prove one of them and refer the interested reader to the aforementioned thesis [69]. Most of the metric results (Khintchine–Groshev, Jarnik, Jarnik–Besicovitch, and Schmidt theorems) stem from the Dirichlet type result which is stated and proved below for the current settings. Throughout, I set $N = \{2^t : t \in \mathbb{N}\}$.

Lemma 1 *For $N_0 < N$, for each $X \in \mathbb{I}^{mn}$ there exists a non-zero integer vector $\mathbf{q}$ in $\mathbb{Z}^m$ and $p \in \mathbb{Z}$ with $|\mathbf{q}|, |p| \leq N$ for N_0 large enough such that*

$$|\mathbf{q}X - p| < (m+2)2N^{-\frac{m+1}{n}+1}$$

Proof: For $|p| < N$ and those $\mathbf{q}$ with non-negative components, there are $(N+1)^m N$ possible vectors of the form $\mathbf{q}X - p$ for which

$$-\frac{m+2}{2}N \leq \mathbf{q}X - p \leq \frac{m+2}{2}N$$

Divide the cube with centre $\mathbf{0}$ and side length $(m+2)N$ in $\mathbb{R}^n$ into N^{m+1} smaller cubes of volume $(m+2)^n N^{n-m-1}$ and side length $(m+2)N^{1-\frac{m+1}{n}}$. Since $N^m < (N+1)^m$, there are at least two vectors $\mathbf{q}_1 X - p_1, \mathbf{q}_2 X - p_2$, say, in one small cube. Therefore

$$\left|(\mathbf{q}_1 - \mathbf{q}_2)X - (p_1 - p_2)\right| < (m+2)2N^{-\frac{m+1}{n}+1}$$

Evidently $\mathbf{q}_1 - \mathbf{q}_2 \in \mathbb{Z}^m$ and $|\mathbf{q}_1 - \mathbf{q}_2| \leq N$. Also, $p_1 - p_2 \in \mathbb{Z}$ and $|p_1 - p_2| \leq N$ by choices of p_1 and p_2.

Lemma 2 *The family $\mathcal{R}_q$ is locally ubiquitous with respect to the function $\rho : \mathbb{N} \to \mathbb{R}^+$ where*

$$\rho(t) = (m+2)2N^{-\frac{m+1}{n}+1}\omega(t)$$

where $\omega(t)$ is a positive real increasing function such that $\omega(t) \to \infty$ as $t \to \infty$. However, it is not very restrictive in the sense that it can always be assumed as a step function and hence does not appear in the sum condition; for details see [69, page 83].

In view of Lemma 1, it is natural to consider the following badly approximable set. Let $\mathrm{Bad}(m,n)$ denote the set of $X \in \mathbb{I}^{mn}$ for which there exists a constant $C(X) > 0$ such that

$$|\mathbf{q}X - p| > C(X)|\mathbf{q}|^{-\frac{m+1}{n}+1} \quad \text{for all} \quad p \times \mathbf{q} \in \mathbb{Z}^{m+1} \tag{4.6}$$

Theorem 9 *Let $m+1 > n$; then*

$$\dim \mathrm{Bad}(m,n) = mn$$

and for $m+1 \leq n$

$$|\mathrm{Bad}(m,n)|_{mn} = 1$$

The proof of Theorem 9 follows from [71, 73] by setting $u = 1$. Now, for $m+1 > n$, since $\mathrm{Bad}(m,n) \subseteq \mathbb{I}^{mn} \setminus W(m,n;\psi)$, therefore $|\mathrm{Bad}(m,n)|_{mn} = 0$.

Remark 1 *It should be clear from Theorem 9 that the minimum distance between* $\mathbf{q}X$ *and the nearest integer vector* $(p, \cdots, p)$ *is at least* $C(X)|\mathbf{q}|^{-\frac{m+1}{n}+1}$*, where* $C(X) > 0$ *is a constant. Loosely speaking,* $\mathrm{Bad}(m,n)$ *consists of all those points that stay clear of* $(m-1)n$*-dimensional hyperplanes having diameter proportional to* $|\mathbf{q}|^{-\frac{m+1}{n}+1}$ *centered at the hyperplanes* R_q*. Note that if the exponent* $-\frac{m+1}{n}+1$ *is replaced by* $-\frac{m+1}{n}+1-\epsilon$ *for* $\epsilon > 0$*, then the set* $\mathrm{Bad}(m,n)$ *is a full Lebesgue measure. It is very pleasing and aligned with our applications.*

Remark 2 *In the case* $m+1 \leq n$*, the set* $W(m,n;\psi)$ *is over determined and lies in a subset of strictly lower dimension than mn.*

To see this consider the case $m = n$ *and* $\det X \neq 0$*. This would imply that the defining inequalities* (4.4) *take the form*

$$|\mathbf{q} - \mathbf{p}X^{-1}| \leq C(X)\psi(|\mathbf{q}|)$$

which is obviously not true for sufficiently large $\mathbf{q}$.

The same logic extends to all other cases. For each $m \times n$ *matrix* $X \in \mathbb{R}^{mn}$ *with column vectors* $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}$ *define* $\tilde{X}$ *to be the* $m \times (n-1)$ *matrix with column vectors* $\mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(n)}$*. The set* $\Gamma \subset \mathbb{R}^{mn}$ *is the set of* $X \in \mathbb{R}^{mn}$ *such that the determinant of each* $m \times m$ *minor of* $\tilde{X}$ *is zero.*

It will now be proved that $W(m,n;\psi) \subset \Gamma$ *when* $m+1 \leq n$*.*

Lemma 3 *For* $m+1 \leq n$ *the set* $W(m,n;\psi)$ *is contained in* Γ*, and* $\dim\Gamma = (m-1)n + m < mn$*. Thus,*

$$\dim W(m,n;\psi) \leq (m-1)n + m$$

The proof of the Lemma can be easily adapted from [69], and the related metric theory can also be proved similarly.

4.5 DoF of $(K\times 2, M)$ X Channel with Constant Real Channel Gains

Using the Khintchine-Groshev theorem is the main tool in calculating the total DoF of the single antenna X channel [24]. This brilliant theorem cannot be applied to MIMO systems. In this case, it is required to extend the Khintchine–Groshev theorem to linear forms over vectors. It was expected that such extension can be solved under simultaneous Diophantine approximation. It is observed that a modification to such an extension gives us a powerful tool, but it cannot be used directly for MIMO X channels. Interestingly, using a similar approach to the original proof of the Khintchine–Groshev theorem for the vectors made it possible to provide a similar type of results for the badly approximable number sets, leading to a new Diophantine approximation theorem. This technique empowers us to deploy the Khintchine–Groshev type of theorems for the MIMO X channel. This tool enables the power of joint processing/simultaneous decoding at all received antennas of a particular receiver, which is crucial in approximation of sets of numbers in higher dimensions and consequently decoding in the MIMO setup.

Let us assume that the X channel consists of K transmitters and two receivers. Each node in the channel is equipped with M antennas.

4.5.1 Encoding

The ith transmitter sends two sets of messages, $\boldsymbol{u}^i = (u^i_1, u^i_2, ...u^i_M)^t$ and $\boldsymbol{v}^i = (v^i_1, v^i_2, ...v^i_M)^t$. It is preferable to decode these at receivers 1 and 2, respectively. The transmitter selects its modulation points from $\mathcal{U} = A(-Q, Q)_{\mathbb{Z}}$ and $\mathcal{V} = A(-Q, Q)_{\mathbb{Z}}$ for u^i_l and v^i_l, $l = 1, 2, .., M$, accordingly. A is a constant factor that controls the minimum distance of the received constellation.

The transmit directions are first chosen in such a way that the interfering signals at both receivers are aligned, i.e., they arrive at the receiver with the same coefficient. To this end, two $M \times M$ matrices $\mathbf{I}^1$ and $\mathbf{I}^2$ are fixed at receivers 1 and 2, respectively, and these coefficients are dedicated to the interfering signals. $\mathbf{I}^1$ and $\mathbf{I}^2$ can be used to design the transmitted signals from all transmitters. For instance, the ith transmitter uses the following signal for data transmission:

$$\boldsymbol{x}^i = (\mathbf{H}^{i,2})^{-1}\mathbf{I}^2\boldsymbol{u}^i + (\mathbf{H}^{i,1})^{-1}\mathbf{I}^1\boldsymbol{v}^i \tag{4.7}$$

4.5.2 Decoding

Using the preceding signaling scheme, the received signals are

$$\begin{cases} \boldsymbol{y}^1 = \sum_{i=1}^{K}(\mathbf{H}^{i,1})(\mathbf{H}^{i,2})^{-1}\mathbf{I}^2\boldsymbol{u}^i + \mathbf{I}^1\sum_{i=1}^{K}\boldsymbol{v}^i + \boldsymbol{z}^1 \\ \boldsymbol{y}^2 = \sum_{i=1}^{K}(\mathbf{H}^{i,2})(\mathbf{H}^{i,1})^{-1}\mathbf{I}^1\boldsymbol{v}^i + \mathbf{I}^2\sum_{i=1}^{K}\boldsymbol{u}^i + \boldsymbol{z}^2 \end{cases} \tag{4.8}$$

where $\boldsymbol{z}^1$ and $\boldsymbol{z}^2$ are independent Gaussian random vectors with identity covariance matrices. At the lth antenna of the first receiver,

$$y_l^1 = \sum_{i=1,\dots,K}\sum_{j=1,\dots,M} g_{l,j}^i u_j^i + \sum_{j=1,\dots,M}\eta_{lj}\Gamma_j + z_l^1 \tag{4.9}$$

where $g_{l,j}^i$ is the received gain (coefficient) for each u_j^i observed at the lth antenna. This number puts together all the effects of pre-coding and channel gain; $\eta_{i,j}$ is the ith row, jth column component of matrix $\mathbf{I}^1$ and $\Gamma_j=\sum_{i=1}^{K}v_j^i$. Similarly, at the lth antenna of the second receiver

$$y_l^2 = \sum_{i=1,\dots,K}\sum_{j=1,\dots,M} \hat{g}_{l,j}^i v_j^i + \sum_{j=1,\dots,M}\lambda_{l,j}\Theta_j + z_l^2 \tag{4.10}$$

where $\mathbf{I}^2=[\lambda_{i,j}]$ and $\Theta_j=\sum_{i=1}^{k}u_j^i$.

How the first receiver decodes a given desired message, say u_1^1, from the received signals from all antennas is now described. One can normalize the received signal to set the coefficients of u_1^1 at all antennas equal to unity. Next, the joint processing method is applied to decode u_1^1 simultaneously from all antennas. Theorem 8 allows the minimum distance to be approximated by $d_{\min}=AQ^{-k}$. Hence, setting $A \approx Q^k$ is sufficient to $d_{\min} \approx 1$, which in turn results in noise removal from the received signal. Putting this together gives $P \approx Q^{2(k+1)}$. At the first receiver the following DoF per each favorite signal is obtained (this technique can be applied to all u_{ij} at the first receiver):

$$d^{1,1} = \lim_{P\to\infty}\frac{(R^{1,1} = \log(2Q-1))}{0.5\log P} = \frac{1}{K+1} \tag{4.11}$$

In the second receiver, the same method will be applied for all v^i_j, resulting in the same DoF for the second receiver as well. Finally, at each receiver it is possible to decode KM different messages, which results in the total DoF of $\frac{2KM}{K+1}$ for the system. This, in fact, meets the upper bound mentioned in [23].

4.6 DoF of $(2\times K, M)$ X Channel with Constant Real Channel Gains

Interestingly, layered interference alignment can be deployed to show that the reciprocity along with duality are held for $K \times 2$, MIMO X channels. In the following we will show that the total DoF of $(2 \times K, M)$ antenna X channel with constant real channel gain realization is the same as the DoF of $(K \times 2, M)$ X channel and is $\frac{2KM}{K+1}$.

4.6.1 Encoding

The first transmitter sends K sets of messages, $\boldsymbol{u}^j = (u^j_1, u^j_2, ...u^j_M)^t$ for $j = 1, .., K$. The second transmitter similarly transmits another K sets of messages, $\boldsymbol{v}^j = (v^j_1, v^j_2, ...v^j_M)^t$; it is preferred that $\boldsymbol{u}^j$ and $\boldsymbol{v}^j$ are decoded at receiver j. The transmitter selects its modulation points from $\mathcal{U} = A(-Q, Q)_{\mathbb{Z}}$ and $\mathcal{V} = A(-Q, Q)_{\mathbb{Z}}$ for u^j_l and v^j_l, $l = 1, 2, ..., M$, respectively. A is a constant factor that controls the minimum distance of the received constellation.

Like $(K \times 2,\ M)$ X channel, the transmit directions are first chosen in such a way that the interfering signals at both receivers are aligned. To this end, matrices $\mathbf{I}^i$ are fixed at receiver i, each of dimension $M \times M$. $\mathbf{I}^i$'s can be used to obtain the transmit signals from all transmitters. The desired goal for each receiver i is

$$\boldsymbol{y}^i = \mathbf{H}^{1,i}\rho^i\boldsymbol{u}^i + \mathbf{H}^{2,i}\zeta^i\boldsymbol{v}^i + \sum_{j=1\&i\neq j}^{K} \mathbf{I}^i + \boldsymbol{z}^i \tag{4.12}$$

To obtain ρ and ζ values, the following solution is proposed:

$$\begin{cases} \mathbf{H}^{1,i}\rho^j = \mathbf{H}^{2,i}\zeta^{j+1} & j \notin \{i, i-1, K\} \\ \mathbf{H}^{1,i}\rho^j = \mathbf{H}^{2,i}\zeta^{j+2} & j = i-1 \\ \mathbf{H}^{1,i}\rho^j = \mathbf{H}^{2,i}\zeta^{1} & j = K \;\&\; i \neq 1 \\ \mathbf{H}^{1,i}\rho^j = \mathbf{H}^{2,i}\zeta^{2} & j = K \;\&\; i = 1 \end{cases}$$

Using the above signal space design gives

$$\mathbf{I}^j = \begin{cases} \mathbf{H}^{1,i}\rho^j(\boldsymbol{u}^j + \boldsymbol{v}^{j+1}) & j \notin \{i, i-1, K\} \\ \mathbf{H}^{1,i}\rho^j(\boldsymbol{u}^j + \boldsymbol{v}^{j+2}) & j = i-1 \\ \mathbf{H}^{1,i}\rho^j(\boldsymbol{u}^j + \boldsymbol{v}^{1}) & j = K \;\&\; i \neq 1 \\ \mathbf{H}^{1,i}\rho^j(\boldsymbol{u}^j + \boldsymbol{v}^{2}) & j = K \;\&\; i = 1 \end{cases}$$

4.6.2 Decoding

With the preceding signaling scheme, the received signals at the lth antenna of the receiver j can be modelled as

$$y_l^j = \sum_{i=1,\ldots,M} \sigma_{l,i} u_j^i + \sum_{i=1,\ldots,M} \lambda_{l,i} v_j^i + \sum_{i=1,i\neq j}^{K} \sum_{n=1}^{M} I_n^i + z_l^j \tag{4.13}$$

where $\sigma_{l,i}$ and $\lambda_{l,i}$ are constant coefficients representing the product of all the gain effects for the received u_j^i and v_j^i, respectively.

Now, applying the joint processing technique at each antenna, we have received the linear combination of $2M$ desired message parts (M for $\boldsymbol{u}$ and M for $\boldsymbol{v}$) added to $M(K-1)$ non-desired interference terms. For any message u_j^i at the lth antenna of the receiver j using joint processing among all the M antennas, after normalizing and deploying the results of Theorem 8, it can be concluded that

$$d^{i,j} = \lim_{P\to\infty} \frac{\log(2Q-1)}{0.5\log P} = \frac{1}{K+1} \tag{4.14}$$

The same argument can be put forward for v_j^i, so it is concluded that the total DoF will be $\frac{2KM}{K+1}$. It is observed that $(2\times K, M)$ and $(K\times 2, M)$ X channels act reciprocal/dual as long as the DoF is the merit. Here, for both $(2\times K, M)$ and $(K\times 2, M)$ X channels the achievability part is proved, since in [23] it is shown that the total DoF for both $(2\times K, M)$ and $(K\times 2, M)$ X channels are bounded by $\frac{2KM}{K+1}$. Therefore, it can be concluded that the achievable scheme described here has characterized the total DoF of these channels.

4.7 Complex Channel Coefficients

Let us consider the $(K\times 2, M)$ X channel. The upper bound on the total DoF of $\frac{2KM}{K+1}$ is achievable for this channel when the channel gains are real numbers. Earlier, I have proved this for both constant and time varying real channel gains.

Needless to say, the result is also applicable to channels with complex coefficients. The real and imaginary parts of the input and the output can be paired. This converts the channel to $2K$ virtual transmitters and 4 receivers. Using Theorem 8, it can be seen that the total DoF of this channel cannot be achieved by an application of that theorem.

Interestingly, it was possible to prove that there can be an extension for most of the complex channel realization to Theorem 8, which can be used in characterizing the total DoF of both $(K \times 2, M)$ and $(2 \times K, M)$ X channels with complex constant channel gains. The same signal design at the encoder is considered along with similar joint decoding at the receiver side. This means the layered interference alignment can almost surely characterize these channels DoF as long as the channel gains are constant numbers.

The details and proof of the new theorem are discussed in detail in section 4.8.1. Using Theorem 13 shows that the total achievable DoF using the traditional definition of it will be twice that of the case of constant real channel gains. This chain is acceptable and can be justified with either the fact that for a complex channel two dimensions per transmission (real and imaginary) are needed or by using the definition of DoF for the case of complex channel gains and complex number transmission, which is

$$d^i = \lim_{P\to\infty} \frac{R^i}{logP} \tag{4.15}$$

in which R^i is the maximum rate of reliable ith transmission. Using the above definition shows that the total DoF of these channels are the same as $\frac{2KM}{K+1}$.

It is worth noting that joint processing between all antennas and/or real-imaginary parts at a transmitter increases the achievable sum rate of the channel. This plays a major role in collaboration with the joint processing/simultaneous decoding, even at high SNR regimes. However, in [24], for the GIC it is observed that the total DoF of the channel, which is defined at high SNR regimes, the increase vanishes and the total DoF of the channel can be achieved by separate coding over all available dimensions.

4.8 Metric Diophantine Approximation over Complex Numbers

In the 19th century, Hermite and Hurwitz studied the approximation of complex numbers by the ratios of Gaussian integers, a natural analogue of approximation of real numbers by rationals,

$$\mathbb{Z}[i] = \{p_1 + ip_2 \in \mathbb{C} : p_1, p_2 \in \mathbb{Z}\}$$

However, complex Diophantine approximation appears to be more difficult than the real case. For example, continued fractions, so simple and effective for real numbers, are not so straightforward for complex numbers. In other words, by means of continued fraction expansion approach the best possible analogue of Dirichlet's theorem cannot be derived.

4.8.1 General Setup

The problem for the linear form setup is now discussed, and the recent developments so far for the particular cases will be listed. Let Ψ be an *approximating* function. An $m \times n$ matrix

$$Z = (z_{ij}) := \begin{pmatrix} z_{11} & \cdots & z_{1n} \\ z_{21} & \cdots & z_{2n} \\ \vdots & & \vdots \\ z_{m1} & \cdots & z_{mn} \end{pmatrix} \in \mathbb{C}^{mn}$$

is said to be Ψ-approximable if the system of inequalities

$$|q_1 z_{1j} + q_2 z_{2j} + \cdots + q_m z_{mi} - p_j| < \Psi(|\mathbf{q}|_2) \quad (1 \leq j \leq n) \tag{4.16}$$

is satisfied for infinitely many vectors $\mathbf{p} \times \mathbf{q} \in \mathbb{Z}^n[i] \times \mathbb{Z}^m[i] \setminus \{\mathbf{0}\}$. Throughout, the system (4.16) will be written more concisely as $\mathbf{q}Z$. Here $|\mathbf{q}|_2 = \max\{|q_1|_2, \cdots, |q_m|_2\}$, where for $q_k = q_{k_1} + iq_{k_2} \in \mathbb{Z}[i]$, $|q_k|_2 = \sqrt{|q_{k_1}|^2 + |q_{k_2}|^2}$.

As in the real case, the stemming point of such approximation properties is the Dirichlet theorem. A short and more direct geometry of numbers proof of the complex version of Dirichlet's theorem is given below. Although the constant here is not the best possible, the result is all that is needed to prove the complex analogue of Khintchine–Groshev and Schmidt type theorems without recourse to the hyperbolic space framework.

Theorem 10 *Given any $Z \in \mathbb{C}^{mn}$ and $N \in \mathbb{N}$, there exist Gaussian integers $\mathbf{p} = (p_{11} + ip_{12}, \cdots, p_{n1}+ip_{n2}) \in \mathbb{Z}^n[i]$ and non-zero $\mathbf{q} = (q_{11}+iq_{12}, \cdots, q_{m1}+iq_{m2}) \in \mathbb{Z}^m[i]$ with $0 < |\mathbf{q}| \leq N$ such that*

$$|\mathbf{q}Z - \mathbf{p}| < \frac{c}{N^{m/n}} \tag{4.17}$$

where $c > 0$ is an appropriate constant. Moreover, there are infinitely many $\mathbf{p}, \mathbf{q} \in \mathbb{Z}^n[i]\times\mathbb{Z}^m[i]\setminus \{\mathbf{0}\}$ such that

$$|\mathbf{q}Z - \mathbf{p}| < \frac{c}{|\mathbf{q}|_2^{m/n}} \tag{4.18}$$

Proof: For clarity, the theorem for $m = 2, n = 1$ is provided. The proof of the case $m = n = 1$ can be found in [74]. Let $Z = (x_1 + iy_1, x_2 + iy_2), \mathbf{q} = (q_{11} + iq_{12}, q_{21} + iq_{22})$, and $p = (p_1 + ip_2)$. Then

$$|\mathbf{q}Z - \mathbf{p}| = |q_{11}x_1 + q_{21}x_2 - q_{12}y_1 - q_{22}y_2 - p_1 + i(q_{12}x_1 + q_{22}x_2 + q_{11}y_1 + q_{21}y_2 - p_2)| \tag{4.19}$$

Consider the convex body

$$B = \left\{(q_{11}, q_{12}, q_{21}, q_{22}, p_1, p_2) : \max\{q_{11}^2 + q_{12}^2, q_{21}^2 + q_{22}^2\} \leq N^2, \Delta \leq R^2\right\} \tag{4.20}$$

where

$$\Delta = (q_{11}x_1 + q_{21}x_2 - q_{12}y_1 - q_{22}y_2 - p_1)^2 + (q_{12}x_1 + q_{22}x_2 + q_{11}y_1 + q_{21}y_2 - p_2)^2 \tag{4.21}$$

Then

$$\begin{aligned}|B| &= \int_{\max\{q_{11}^2+q_{12}^2,\ q_{21}^2+q_{22}^2\}\le N^2} \int_{\Delta\le R^2} dq_{11}dq_{12}dq_{21}dq_{22}dp_1dp_2 \\ &= \int_{\max\{q_{11}^2+q_{12}^2,\ q_{21}^2+q_{22}^2\}\le N^2} \pi R^2 dq_{11}dq_{12}dq_{21}dq_{22} \\ &= \pi^3 R^2 N^4 \ge 2^6\end{aligned}$$

if $R > \frac{2^3}{\pi^{3/4}N^2}$. Hence, by Minkowski's theorem [75], (4.17) has a non-zero integer solution with $0 < |\mathbf{q}| \le N$.

This result should be compared with the real Dirichlet's theorem in Section 3.5 for $m = 4, n = 2$. The complex points for which Theorem 10 cannot be improved by an arbitrary constant are called badly approximable. The point $Z \in \mathbb{C}^{mn}$ is said to be badly approximable if there exists a constant $C(Z) > 0$ such that

$$|\mathbf{q}Z - \mathbf{p}| > C(Z)|\mathbf{q}|_2^{-\frac{m}{n}}$$

for all $\mathbf{p} \times \mathbf{q} \in \mathbb{Z}^n[i] \times \mathbb{Z}^m[i]$. Let $\mathbf{Bad}_{\mathbb{C}}(m,n)$ denote the set of badly approximable points in $\mathbb{C}^{mn}$.

The Hausdorff dimension of the set $\mathbf{Bad}_{\mathbb{C}}(1,1)$ has been studied by various authors in different frameworks; see, for instance, [76, §5.3] in which the authors determined the Hausdorff dimension for $\mathbf{Bad}_{\mathbb{C}}(1,n)$, i.e.,

$$\dim \mathbf{Bad}_{\mathbb{C}}(1,n) = n$$

In fact, as a consequence of the general framework in their thesis they proved the weighted analogue intersected with any compact subset of $\mathbb{C}^n$. Now the problem here is whether it is possible to prove the Hausdorff dimension for $\mathbf{Bad}_{\mathbb{C}}(m,n)$.

Problem 1 1 *Can we prove or disprove*

$$\dim \boldsymbol{Bad}_{\mathbb{C}}(m,n) = mn$$

From now onwards we restrict ourselves to the mn-dimensional unit disc $D := (\mathbb{C} \cap \Omega)^{mn}$ where $\Omega = \{a + ib : 0 \le a, b < 1\}$ instead of considering the full space $\mathbb{C}^{mn}$. The reason behind this restriction is that it is convenient to work in the unit discs, and the approximable properties (well and bad both) are invariant under the translation by the Gaussian integers. Let $W_{\mathbb{C}}(m,n;\Psi)$ denote the set of Ψ-approximable points in D,

$$W_{\mathbb{C}}(m,n;\Psi) := \{Z \in D : |\mathbf{q}Z - \mathbf{p}| < \Psi(|\mathbf{q}|_2) \text{ for i.m. } (\mathbf{p},\mathbf{q}) \in \mathbb{Z}^n[i] \times \mathbb{Z}^m[i] \setminus \{\mathbf{0}\}\}$$

4.8.2 Khintchine–Groshev Theorem

The aim here is to prove the complex version of the Khintchine–Groshev theorem

Theorem 11 *Let Ψ be an approximating function. Then*

$$|W_{\mathbb{C}}(m,n;\Psi)|_{mn} = \begin{cases} 0 & \text{if} \quad \sum_{r=1}^{\infty} r^{2m-1}\Psi(r)^{2n} < \infty \\ \textit{Full} & \text{if} \quad \sum_{r=1}^{\infty} r^{2m-1}\Psi(r)^{2n} = \infty \end{cases}$$

For $m = n = 1$, Theorem 13 was proved in 1952 by LeVeque [77], who combined Khintchine's continued fraction approach with ideas from hyperbolic geometry. In 1982, Sullivan [78] used Bianchi groups and some powerful hyperbolic geometry arguments to prove more general Khintchine theorems for real and for complex numbers. In the latter case, the result includes approximation of complex numbers by ratios p/q of integers p, q from the imaginary quadratic fields $\mathbb{R}(i\sqrt{d})$, where d is a square-free natural number. The case $d = 1$ corresponds to the Picard group and approximation by Gaussian rationals. The result was also derived by Beresnevich et al. as a consequence of ubiquity framework in [52, Theorem 7].

4.8.3 Proof of the Convergence Case of Theorem 13

As before, the Theorem 13 is proved for the case $m = 2, n = 1$, leaving behind the obvious modifications to deal with the higher dimensions. First, the convergence case is dealt with. The resonant set is defined as

$$\begin{aligned} C_{\mathbf{q}} &:= \{Z \in D : |\mathbf{q}Z - \mathbf{p}| = 0\} \\ &= \{(x_1 + iy_1, x_2 + iy_2) \in D : |(q_{11} + iq_{12}, q_{21} + iq_{22}) \cdot (x_1 + iy_1, x_2 + iy_2) - (p_1 + ip_2)| = 0\} \\ &= \left\{ (x_1 + iy_1, x_2 + iy_2) \in D : \begin{array}{l} q_{11}x_1 + q_{21}x_2 - q_{12}y_1 - q_{22}y_2 = p_1 \text{ and} \\ q_{12}x_1 + q_{22}x_2 + q_{11}y_1 + q_{21}y_2 = p_2 \end{array} \right\} \end{aligned}$$

The set $W_{\mathbb{C}}(2, 1; \Psi)$ can be written as using the resonant sets

$$W_{\mathbb{C}}(2, 1; \Psi) = \bigcap_{N=1}^{\infty} \bigcup_{r>N} \bigcup_{C_{\mathbf{q}}:|\mathbf{p}|_2<|\mathbf{q}|_2=r} B\left(C_{\mathbf{q}}, \Psi(|\mathbf{q}|_2)\right)$$

where

$$B\left(C_{\mathbf{q}}, \Psi(|\mathbf{q}|_2)\right) = \left\{Z \in D : \operatorname{dist}\left(Z, C_{\mathbf{q}}\right) \leq \frac{\Psi(|\mathbf{q}|_2)}{|\mathbf{q}|_2}\right\}$$

It follows that

$$W_{\mathbb{C}}(2, 1; \Psi) \subseteq \bigcup_{r>N} \bigcup_{C_{\mathbf{q}}:|\mathbf{p}|_2<|\mathbf{q}|_2=r} B\left(C_{\mathbf{q}}, \Psi(|\mathbf{q}|_2)\right)$$

In other words, $W_{\mathbb{C}}(2, 1; \Psi)$ has a natural cover $C = \left\{B\left(C_{\mathbf{q}}, \Psi(|\mathbf{q}|_2)\right) : |\mathbf{q}|_2 > N\right\}$ for each $N = 1, 2, \cdots$. This can further be covered by a collection of 4-dimensional hypercubes with disjoint interior and side length comparable with $\Psi(|\mathbf{q}|_2)/|\mathbf{q}|_2$. The number of such hypercubes is clearly $<< (\Psi(|\mathbf{q}|_2)/|\mathbf{q}|_2)^{-2}$. Thus,

$$\begin{aligned} |W_{\mathbb{C}}(2, 1; \Psi)|_2 &\leq \sum_{r=N}^{\infty} \sum_{C_{\mathbf{q}}:|\mathbf{p}|_2<|\mathbf{q}|_2=r}^{\infty} \left|B\left(C_{\mathbf{q}}, \Psi(|\mathbf{q}|_2)\right)\right|_2 \\ &\ll \sum_{r=N}^{\infty} \sum_{r<|\mathbf{q}|_2<r+1} (\Psi(|\mathbf{q}|_2)/|\mathbf{q}|_2)^{-2} (\Psi(|\mathbf{q}|_2)/|\mathbf{q}|_2)^4 \\ &= \sum_{r=N}^{\infty} (\Psi(r)/r)^2 \sum_{r<|\mathbf{q}|_2<r+1} 1 \end{aligned} \tag{4.22}$$

Now it remains to count $\sum\limits_{r<|\mathbf{q}|_2<r+1} 1$. An argument from [74, p. 328] or [75, Th. 386] is followed to conclude that $\sum_{r<|\mathbf{q}|_2<r+1} 1 \ll r^5$. Thus, (4.22) becomes

$$|W_{\mathbb{C}}(2, 1; \Psi)|_2 \ll \sum_{r=N}^{\infty} r^3 \Psi(r)^2$$

Now, since the sum $\sum_{r=N}^{\infty} r^3\Psi(r)^2 < \infty$, the tail of the series can be made arbitrarily small. Hence, by the Borel–Cantelli lemma, $|W_{\mathbb{C}}(2,1;\Psi)|_2 = 0$.

The divergence case of the above theorem can be similarly proved by following the similar arguments as in the real case. Precisely, one would need to utilize the ubiquity framework to extend [52, Th. 7] for the linear forms setup. The Dirichlet theorem 10 would again be used to prove the ubiquity lemma. The details are left to the interested reader.

4.8.4 A Complex Hybrid Set

As in the previous section, let Ψ be an *approximating* function. An $m \times n$ matrix $Z \in \mathbb{C}^{mn}$ is said to be Ψ-approximable if the system of inequalities

$$|q_1 z_{1j} + q_2 z_{2j} + \cdots + q_m z_{mi} - p| < \Psi(|\mathbf{q}|_2) \quad (1 \leq j \leq n) \tag{4.23}$$

is satisfied for infinitely many vectors $(p, \cdots, p, q_1, \cdots, q_m) \in \mathbb{Z}^n[i] \times \mathbb{Z}^m[i] \setminus \{\mathbf{0}\}$. That is, the system (4.23) is obtained by keeping the nearest integer vector $(p, \cdots, p)$ the same for all the linear forms. Since the result are very similar in nature to that of the classical sets $W_A(m, n; \Psi)$ and can be proved analogously, they are only stated here. The first one is the Dirichlet type theorem, and rest of the results stems from it. It also serves the purpose of finding the minimum distance between $\mathbf{q}Z$ and $\mathbf{p}$.

Theorem 12 *Given any $Z \in \mathbb{C}^{mn}$ and $N \in \mathbb{N}$, there exist Gaussian integers $\mathbf{p} = (p_1+ip_2, \cdots, p_1+ip_2) \in \mathbb{Z}^n[i]$ and non-zero $\mathbf{q} = (q_{11} + iq_{12}, \cdots, q_{m1} + iq_{m2}) \in \mathbb{Z}^m[i]$ with $0 < |\mathbf{q}| \leq N$ such that*

$$|\mathbf{q}Z - \mathbf{p}| < \frac{c}{N^{\frac{m+1}{n}-1}} \tag{4.24}$$

where $c > 0$ is an appropriate constant. Moreover, there are infinitely many $(\mathbf{p}, \mathbf{q}) \in \mathbb{Z}^n[i] \times \mathbb{Z}^m[i] \setminus \{\mathbf{0}\}$ such that

$$|\mathbf{q}Z - \mathbf{p}| < c|\mathbf{q}|_2^{-\frac{m+1}{n}+1} \tag{4.25}$$

Let $W_{\mathbb{C}_0}(m, n; \Psi)$ denote the set of Ψ-approximable points in D, i.e., the set of points that satisfy the system (4.23). Then, one has the analogue of the Khintchine–Groshev theorem for this setup.

Theorem 13 *Let Ψ be an approximating function and let $m + 1 > n$. Then*

$$|W_{\mathbb{C}_0}(m, n; \Psi)|_{mn} = \begin{cases} 0 & \text{if} \quad \sum_{r=1}^{\infty} (r^{m-n}\Psi(r)^n)^2 < \infty \\ Full & \text{if} \quad \sum_{r=1}^{\infty} (r^{m-n}\Psi(r)^n)^2 = \infty \end{cases}$$

The proof of this theorem is again similar to Theorem 13. The details are left to the interested reader.

Chapter 5

Conclusion and Future Research Directions

5.1 Conclusion

In this thesis, a new interference management tool named the *Layered Interference Alignment* was introduced. For the first time a powerful theorem in the field of number theory was introduced and proved for both real and complex numbers. This number theoretic theorem empowers us to deploy multiple antenna collaboration, which leads to simultaneous decoding and joint processing. It is observed that despite GIC and SISO X channel, joint processing is required to characterize the total DoF of a class MIMO X channel. To this end both the vector and the real interference alignment techniques for signal transmission were incorporated and a joint processing scheme for simultaneous decoding exploited. Eventually, the total DoF of $(K \times 2, M)$ and $(2 \times K, M)$ X channels are characterized. It is observed that regardless of complex or real channel realization for constant channel gains both of these channels can achieve the total DoF outerbound of $\frac{2KM}{K+1}$.

This new technique relies on the rational dimension and needs a complicated encoding and simultaneous decoding, but it has less sensitivity to the rationality of numbers than its brother, the real interference alignment. This is because the total dimensionality of space is comprised of higher dimensional space, and its trajectory over each physical dimension (span of the space defined by antennas) consists of up to infinite possible rational dimensions. In other words, the coding design is benefiting from the existence of physical dimensions along with the possibility of having infinite rational dimensions over each physical dimension. To be able to design a decoder for this system one must incorporate the entire received streams in all physical dimensions and use joint decoding among all the antennas of each receiver. It was anticipated that this technique would be useful, but it was not in use because of the lack of any existing mathematical theorem supporting the design of a hard decoder at a high SNR regime. For the first time, in this dissertation, the author settled the required mathematical theorem for the real numbers and then extended the theorem over the field of complex numbers. These new Diophantine approximation theorems are a really powerful tool that can bring co-existence of both previously known interference alignment techniques along with traditional multiplexing gain methods of MIMO networks.

5.2 Future Research Directions

Some interesting problems that emerge from this dissertation are discussed here. These problems can provide the spur to further research.

Interference Alignment

With the advent of interference alignment, new directions in interference management came into existence as interference alignment emerged as a promising method to mitigate the effect of interference in a network. The major drawback regarding interference alignment is that it needs full channel state information to realize its full potential. Therefore, practical applications are possible only when efficient feedback strategies are designed and carefully analyzed.

Interference Alignment and Secrecy

Providing a secure communication over networks will be of fundamental importance in the future. A secure system can be obtained by sacrificing available resources. However, as the resources are scarce, this results in tremendous loss in the throughput of the system. Interference alignment, though, can be used in a different fashion to provide security and performance at the same time. At present interference caused by several users can be accumulated for the eavesdropper, and in the future it could be aligned for the intended users to increase the available DoF.

Finite Precision for Rational Dimensions

Both the layered and the real interference alignment methods are known to be powerful techniques to establish asymptotic results like DoF characterization. However, it is not clear whether these techniques can predict the channel capacity in a finite SNR regime. As mentioned throughout this dissertation, the DoF of the K-user constant IC and X channels are discontinuous functions of channel coefficients and are sensitive to the rationality/irrationality of channel coefficients. Specifically, one might argue that the irrationally of the channel coefficients is fundamental in both layered and real interference alignment, and hence the scheme might not work in the presence of unavoidable quantization errors. Recently, it was shown in [79] that the real interference alignment can be used to obtain constant gap capacity results for the two-user X channel. This important study proved that, at least for the two-user X channel, the everywhere discontinuity of the DoF in the channel coefficients is indeed a consequence of the definition of DoF as a limiting expression and not fundamental to the real interference alignment. An interesting future direction is to combine the extended version of the layered interference alignment described here with the method developed in [79] to obtain constant gap capacity characterization for the general MIMO X channel and GIC.

K-user Interference Channels

The coding scheme used for the two- and three-user with rational coefficient case can be brought to the K-user Gaussian IC. In fact, in wireless systems channel estimation is always performed with finite precision, and therefore it is rational. Hence, in the case of three users, a careful design is needed to achieve higher multiplexing gains in the channel. It is also interesting to obtain the relation between the channel coefficients and achievable DOF. Recently, the detailed coding scheme was presented for the three-user MIMO GIC, which brings hope for an immediate solution for the generalized K-user case.

Bibliography

[1] C. Shannon, “Two-way communication channels,” in *Proc. 4th Berkeley Symp. Math. Stat. Prob*, vol. 1, pp. 351–384, 1961.

[2] R. Ahlswede, “Multi-way communication channels,” in *2nd International Symposium on Information Theory, Tsahkadsor, Armenia, USSR, September 2-8, 1971*, p. 23, Akadémiai Kiadó, 1973.

[3] I. Csiszar and J. Korner, *Information theory: coding theorems for discrete memoryless systems*. Academic Press, Inc. Orlando, FL, USA, 1982.

[4] R. Cheng and S. Verdu, “On limiting characterizations of memoryless multiuser capacity regions,” *IEEE Transactions on Information Theory*, vol. 39, no. 2, pp. 609–612, 1993.

[5] T. Cover and J. Thomas, *Elements of information theory*. John Wiley and sons, 2006.

[6] A. Carleial, “A case where interference does not reduce capacity,” *IEEE Trans. Inf. Theory*, vol. 21, p. 569, September 1975.

[7] H. Sato, “The capacity of the Gaussian interference channel under strong interference,” *IEEE Trans. Inf. Theory*, vol. 27, pp. 786–788, November 1981.

[8] A. Motahari and A. Khandani, “Capacity bounds for the Gaussian interference channel,” *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 620–643, 2009.

[9] V. Annapureddy and V. Veeravalli, “Gaussian interference networks: Sum capacity in the low interference regime and new outer bounds on the capacity region,” *submitted to IEEE Trans. Inf. Theory. http://arxiv.org/abs/0802.3495*, 2008.

[10] I. Sason, "On achievable rate regions for the Gaussian interference channel," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1345–1356, 2004.

[11] H. Sato, "On degraded Gaussian two-user channels," *IEEE Trans. Inf. Theory*, vol. 24, no. 5, pp. 637–640, 1978.

[12] M. Costa, "On the Gaussian interference channel," *IEEE Trans. Inf. Theory*, vol. 31, no. 5, pp. 607–615, 1985.

[13] R. Etkin, D. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5534–5562, 2008.

[14] M. Maddah-Ali, A. Motahari, and A. Khandani, "Signaling over mimo multi-base systems: Combination of multi-access and broadcast schemes," in *IEEE Int. Symp. Information Theory*, pp. 2104–2108, Citeseer, 2006.

[15] M. Maddah-Ali, A. Motahari, and A. Khandani, "Communication over MIMO X channels: Interference alignment, decomposition, and performance analysis," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3457–3470, 2008.

[16] "Real interference alignment: Exploiting the potential of single aantenna systems," *Arxiv preprint arxiv:0908.2282*, August 2009.

[17] G. Bresler, A. Parekh, and D. N. Tse, "The approximate capacity of the many-to-one and one-to-many gaussian interference channels," *Information Theory, IEEE Transactions on*, vol. 56, no. 9, pp. 4566–4592, 2010.

[18] V. Cadambe and S. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, pp. 3425–3441, Aug. 2008.

[19] R. Etkin and E. Ordentlich, "The degrees-of-freedom of the K-User Gaussian interference channel is discontinuous at rational channel coefficients," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, 2009.

[20] S. Jafar and S. Shamai, "Degrees of freedom region of the mimo x channel," *Information Theory, IEEE Transactions on*, vol. 54, pp. 151 –170, jan. 2008.

[21] A. Motahari, S. Gharan, and A. Khandani, "Real interference alignment with real numbers," *Submitted to IEEE Trans. Inf. Theory*.

[22] V. R. Cadambe and S. A. Jafar, "Interference alignment and the degrees of freedom for the k user interference channel," *arXiv preprint arXiv:0707.0323*, 2007.

[23] V. R. Cadambe and S. A. Jafar, "Degrees of freedom of wireless x networks," in *Information Theory, 2008. ISIT 2008. IEEE International Symposium on*, pp. 1268–1272, IEEE, 2008.

[24] A. S. Motahari, S. O. Gharan, M.-A. Maddah-Ali, and A. K. Khandani, "Real interference alignment: Exploiting the potential of single antenna systems," *arXiv preprint arXiv:0908.2282*, 2009.

[25] J. W. S. Cassels, *An introduction to Diophantine approximation*. Cambridge Tracts in Mathematics and Mathematical Physics, No. 45, New York: Cambridge University Press, 1957.

[26] A. Ghasemi, A. S. Motahari, and A. K. Khandani, "Interference alignment for the k user mimo interference channel," in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, pp. 360–364, IEEE, 2010.

[27] S. Mahboubi, A. Motahari, and A. Khandani, "Layered interference alignment: Achieving the total dof of mimo x channels," pp. 355 –359, jun. 2010.

[28] I. Telatar, "Capacity of multi-antenna Gaussian channels," *European transactions on telecommunications*, 1999.

[29] S. Jafar and M. Fakhereddin, "Degrees of freedom for the MIMO interference channel," *IEEE Trans. Inf. Theory*, vol. 53, no. 7, p. 2637, 2007.

[30] X. Shang, G. Kramer, and B. Chen, "A New Outer Bound and the Noisy-Interference Sum–Rate Capacity for Gaussian Interference Channels," *IEEE Transactions on Information theory*, vol. 55, no. 2, pp. 689–699, 2009.

[31] T. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE transactions on information theory*, vol. 27, no. 1, pp. 49–60, 1981.

[32] O. Koyluoglu, H. El Gamal, L. Lai, and H. Poor, "Interference alignment for secrecy," *preprint*, 2008.

[33] B. Nazer, M. Gastpar, S. Jafar, and S. Vishwanath, "Ergodic interference alignment," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT 2009), Seoul, Korea*, 2009.

[34] S. Sridharan, A. Jafarian, S. Vishwanath, and S. Jafar, "Capacity of symmetric k-user Gaussian very strong interference channels," in *IEEE Global Telecommunication Conf*, 2008.

[35] S. Sridharan, A. Jafarian, S. Vishwanath, S. Jafar, and S. Shamai, "A layered lattice coding scheme for a class of three user Gaussian interference channels," in *2008 46th Annual Allerton Conference on Communication, Control, and Computing*, pp. 531–538, 2008.

[36] R. Etkin and E. Ordentlich, "On the degrees-of-freedom of the K-user Gaussian interference channel," *available in ArXiv*.

[37] A. Motahari, S. Gharan, and A. Khandani, "On the degrees-of-freedom of the three-user gaussian interfererence channel: The symmetric case," in *IEEE International Symposium on Information Theory*, 2009.

[38] S. Jafar and S. Shamai, "Degrees of freedom region for the MIMO X channel," *Arxiv preprint cs/0607099*, 2006.

[39] A. Lapidoth, "On the high-SNR capacity of noncoherent networks," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3025–3036, 2005.

[40] A. Lapidoth, S. Shamai, and M. Wigger, "On the capacity of fading MIMO broadcast channels with imperfect transmitter side-information," *Arxiv preprint cs/0605079*, 2006.

[41] A. Lapidoth and S. Shamai, "Collapse of degrees of freedom in MIMO broadcast with finite precision CSI," in *Proceedings of 43rd Annual Allerton Conference on Communications, Control and Computing*, 2005.

[42] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Transactions on Information Theory*, vol. 52, no. 11, p. 5045, 2006.

[43] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side information," *IEEE Transactions on Information Theory*, vol. 51, no. 2, 2005.

[44] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *IEEE Transactions on Information Theory*, vol. 52, no. 5, p. 1813, 2006.

[45] V. Beresnevich, "A Groshev type theorem for convergence on manifolds," *Acta Mathematica Hungarica*, vol. 94, no. 1, pp. 99–130, 2002.

[46] A. V. Groshev, "Une theoreme sur les systemes des formes lineaires," *Doklady Akad. Nauk SSSR.*, vol. 19, pp. 151–152, 1938.

[47] W. Schmidt, *Diophantine approximation*. Springer Verlag, 1980.

[48] G. Hardy, E. Wright, D. Heath-Brown, and J. Silverman, *An introduction to the theory of numbers*. Clarendon press Oxford, 1979.

[49] V. I. Bernik and M. M. Dodson, *Metric Diophantine approximation on manifolds*, vol. 137 of *Cambridge Tracts in Mathematics*. Cambridge: Cambridge University Press, 1999.

[50] V. Sprindzhuk, *Metric theory of Diophantine approximations*. Halsted Press, 1979.

[51] A. Khintchine, "Einige Sätze über Kettenbrüche, mit Anwendungen auf die Theorie der Diophantischen Approximationen," *Math. Ann.*, vol. 92, no. 1-2, pp. 115–125, 1924.

[52] V. Beresnevich, D. Dickinson, and S. Velani, "Measure theoretic laws for lim sup sets," *Mem. Amer. Math. Soc.*, vol. 179, no. 846, pp. x+91, 2006.

[53] R. Duffin and A. Schaeffer, "Khintchines problem in metric Diophantine approximation," *Duke Math. J*, vol. 8, pp. 243–255, 1941.

[54] G. Harman, *Metric number theory*, vol. 18 of *London Mathematical Society Monographs. New Series*. New York: The Clarendon Press Oxford University Press, 1998.

[55] V. Beresnevich, V. Bernik, M. Dodson, and S. Velani, "Classical metric Diophantine approximation revisited," in *Analytic number theory*, pp. 38–61, Cambridge: Cambridge Univ. Press, 2009.

[56] V. Jarník., "Diophantische approximationen und hausdorffshes mass," *Mat. Sb.*, vol. 36, pp. 371–382, 1929.

[57] A. Besicovitch, "On the fundamental geometrical properties of linearly measurable plane sets of points," *Mathematische Annalen*, vol. 98, no. 1, pp. 422–464, 1928.

[58] V. Jarník., "Über die simultanen diophantischen approximationen," *Math. Z.*, vol. 33, pp. 505–543, 1931.

[59] A. Khintchine, "Zur metrischen theorie der diophantischen approximationen," *Math. Zeitschr.*

[60] M. M. Dodson, "Geometric and probabilistic ideas in the metric theory of Diophantine approximations," *Uspekhi Mat. Nauk*, vol. 48, no. 5(293), pp. 77–106, 1993.

[61] P. X. Gallagher, "Metric simultaneous diophantine approximation. II," *Mathematika*, vol. 12, pp. 123–127, 1965.

[62] V. Beresnevich and S. Velani, "Classical metric Diophantine approximation revisited: the Khintchine-Groshev theorem," *Int. Math. Res. Not. IMRN*, no. 1, pp. 69–86, 2010.

[63] W. M. Schmidt, “A metrical theorem in geometry of numbers,” *Trans. Amer. Math. Soc.*, vol. 95, pp. 516–529, 1960.

[64] M. M. Dodson, B. P. Rynne, and J. A. G. Vickers, “Diophantine approximation and a lower bound for Hausdorff dimension,” *Mathematika*, vol. 37, no. 1, pp. 59–73, 1990.

[65] D. Dickinson and S. L. Velani, “Hausdorff measure and linear forms,” *J. Reine Angew. Math.*, vol. 490, pp. 1–36, 1997.

[66] V. Beresnevich and S. Velani, “Schmidt’s theorem, Hausdorff measures, and slicing,” *Int. Math. Res. Not.*, pp. Art. ID 48794, 24, 2006.

[67] S. H. Mahboubi, A. S. Motahari, and A. K. Khandani, “Layered interference alignment: Achieving the total dof of mimo x-channels,” in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, pp. 355–359, IEEE, 2010.

[68] M. A. Maddah-Ali, “On the degrees of freedom of the compound miso broadcast channels with finite states,” in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, pp. 2273–2277, IEEE, 2010.

[69] D. Dickinson and M. Hussain, “The metric theory of mixed type linear forms,” *Int. J. of Number Theory*, vol. 9(2), pp. 77–90, 2013.

[70] M. Hussain and J. Levesley, “The metrical theory of simultaneously small linear forms,” *Funct. Approx. Comment. Math.*, vol. 48, no. 2, pp. 167–187, 2013.

[71] M. Hussain and S. Kristensen, “Badly approximable systems of linear forms in absolute value,” *Unif. Dist. Theory*, vol. 8, no. 1, pp. 7–15, 2013.

[72] V. Beresnevich and S. Velani, “Ubiquity and a general logarithm law for geodesics,” in *Dynamical systems and Diophantine approximation*, vol. 19 of *Sémin. Congr.*, pp. 21–36, Paris: Soc. Math. France, 2009.

[73] M. Hussain, “A note on badly approximable linear forms,” *Bull. Aust. Math. Soc.*, vol. 83, no. 2, pp. 262–266, 2011.

[74] M. M. Dodson and S. Kristensen, “Hausdorff dimension and Diophantine approximation,” in *Fractal geometry and applications: a jubilee of Benoît Mandelbrot. Part 1*, vol. 72 of *Proc. Sympos. Pure Math.*, pp. 305–347, Providence, RI: Amer. Math. Soc., 2004.

[75] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*. New York: The Clarendon Press Oxford University Press, fifth ed., 1979.

[76] S. Kristensen, R. Thorn, and S. Velani, “Diophantine approximation and badly approximable sets,” *Adv. Math.*, vol. 203, no. 1, pp. 132–169, 2006.

[77] W. J. LeVeque, “Continued fractions and approximations in $k(i)$. I, II,” *Nederl. Akad. Wetensch. Proc. Ser. A.* **55** = *Indagationes Math.*, vol. 14, pp. 526–535, 536–545, 1952.

[78] D. Sullivan, “Disjoint spheres, approximation by imaginary quadratic numbers, and the logarithm law for geodesics,” *Acta Math.*, vol. 149, no. 3-4, pp. 215–237, 1982.

[79] U. Niesen and M. A. Maddah-Ali, “Interference alignment: From degrees-of-freedom to constant-gap capacity approximations,” in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pp. 2077–2081, IEEE, 2012.